CESifo WORKING PAPERS

6708
2017

October 2017

Heterogeneity in Needs and Negative Marginal Tax Rates

Spencer Bastani, Sören Blomquist, Luca Micheletto

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6708*
Category 1: Public Finance

Heterogeneity in Needs and Negative Marginal Tax Rates

Abstract

This paper highlights the possibility that negative marginal tax rates arise in an intensive-margin optimal income tax model where wages are exogenous and preferences are homogeneous, but where agents differ both in skills (labor market productivity) and their needs for a work-related consumption good.

JEL-Codes: H210.

Keywords: nonlinear income taxation, negative marginal tax rates, heterogeneity in needs, redistribution.

Spencer Bastani
Department of Economics & Statistics
Linnaeus University
Sweden - 35106 Växjö
spencer.bastani@lnu.se

Sören Blomquist
Department of Economics
Uppsala University
Sweden - 75120 Uppsala
soren.blomquist@nek.uu.se

Luca Micheletto
University of Milan
Department of Law "Cesare Beccaria"
Italy - 20122 Milan
luca.micheletto@unimi.it

9th October 2017

1 Introduction

One of the few theoretical results regarding the shape of optimal income structures concerns the sign of marginal tax rates. For quite some time, the literature maintained that optimal marginal income tax rates should be non-negative, following the contributions by Seade (1977, 1982).[1] This result was first challenged by Stiglitz (1982) who showed, in the context of a two-type model, that Pareto efficient taxation requires that (in the standard case where it is the high-skilled self-selection constraint which is binding) the marginal tax rate on the high-skilled agents should be negative, except in the limiting case where the two types of labor are perfect substitutes.[2] In a subsequent and influential contribution, Saez (2002) showed that in the presence of an extensive margin of labor supply, optimal marginal tax rates can be negative, yielding support for an Earned Income Tax Credit (EITC). More recently, Choné and Laroque (2010) demonstrated the possibility of negative marginal tax rate in a model without an extensive margin of labor supply, but where agents have heterogeneous preferences regarding their preference for work.[3]

In this paper we propose a previously unnoticed rationale for negative marginal tax rates. We employ a standard intensive-margin optimal income tax model where agents have *identical* preferences but where we allow for heterogeneity in 'needs' for a work-related good/service, i.e. a good/service that some agents need to purchase in order to work. For illustrative purposes we consider a framework

[1] See also Hellwig (2007) for a recent exposition.

[2] The intuition for the result is that, unless the two types of labor are perfect substitutes, by stimulating the labor supply of high-skilled workers the government can rely to some extent on the general equilibrium incidence of the tax in order to compress the wage distribution.

[3] da Costa and Maestri (2017) provides yet another example of contributions emphasizing the possibility of optimal negative marginal tax rates. In their model the result is obtained by modifying the canonical Mirrleesian model to accommodate the assumption that firms have market power in the labor market.

with 'parents' and 'non-parents' and let child care be the good needed by parents in order to work. For simplicity, we assume that one hour of child care is needed for every hour of work.[4] In contrast to Saez (2002), who explore the rationale for negative marginal tax rates created by fixed costs of work at the *extensive* margin, we highlight that costs of work at the *intensive* margin can be a rationale for negative marginal tax rates.

We show analytically the possibility of negative marginal tax rates in the context of a model with two types of agents and where the social welfare function maximized by the government is of the max-min type. In addition, we present a set of numerical simulations showing that our results are not confined to knife-edge cases and that they generalize to settings with more than two types of agents (where negative marginal tax rates can arise but not necessarily at the top of the income distribution) and with different social welfare functions.

The gist of our argument is as follows. We first demonstrate that in a first-best allocation, where the government observes the skills of workers, redistribution might go from low-income agents to high-income agents. More specifically, we highlight circumstances under which a parent earns more than a non-parent but where the total tax payment of the non-parent is positive whereas the parent receives a transfer. We then characterize the circumstances that make such a first-best allocation not implementable under an anonymous nonlinear income tax schedule, i.e. a tax function that is common to all agents. When the first-best optimum is not implementable under an anonymous income tax schedule, we show that at a second-best optimum, while the labor supply of non-parents remains undistorted, the labor supply of parents is distorted upwards by letting them face a negative marginal tax rate. We also show that there may be cases

[4]The interpretation could also be broader as there are other groups of agents who might face needs constraints in practice, such as middle-aged workers who need to purchase care for their elderly parents in order to work, or workers who have commuting costs.

where, while redistribution goes from high- to low-income earners at a first-best optimum, redistribution goes in the opposite direction, and a negative marginal tax rate arises, at a second-best optimum.

The paper is organized as follows. In section 2 we show the possibility of a negative marginal tax rate at the top under a max-min social welfare function. Section 3 provides numerical examples showing that negative marginal tax rates may be optimal also under different social welfare functions and are not necessarily confined to the top of the income distribution. Finally, section 4 concludes.

2 Negative marginal tax rates under a max-min social welfare function

Consider an economy populated by two groups of agents, parents and non-parents. Both have preferences represented by the utility function

$$U = c - \frac{1}{1+\frac{1}{\beta}} h^{1+\frac{1}{\beta}}, \tag{1}$$

where c denotes consumption, h denotes labor supply, and where β is a positive constant representing the elasticity of labor supply. We further denote the wage rate of parents by w^p (> 0) and the wage rate for non-parents by w^{np} (> 0), and assume that parents need one hour of child care services for each hour of work, with q denoting the hourly market price for child care services. Furthermore, assume that $w^p - q > 0$. Earned income is denoted by Y, where $Y = wh$, and after-tax income is denoted by B. Moreover, we let π be the proportion of parents in the population, with the total population normalized to one.

We start by deriving a condition that guarantees that, at a first-best max-min optimum, parents earn more than non-parents, redistribution goes from low-income- to high-income earners, and the first-best max-min optimum is not implementable with an anonymous nonlinear income tax schedule. We then show in proposition 1 below that at a second-best optimum, parents will face a negative marginal tax rate, whereas the labor supply of non-parents is left undistorted.

In a first-best setting, the problem solved by a max-min government is given by:

$$\max_{Y^p, B^p, Y^{np}, B^{np}} \quad B^p - \frac{q}{w^p} Y^p - \frac{1}{1+\frac{1}{\beta}} \left(\frac{Y^p}{w^p} \right)^{1+\frac{1}{\beta}}$$

subject to:

$$\begin{aligned} \pi \left[Y^p - B^p \right] + (1-\pi) \left[Y^{np} - B^{np} \right] &= 0, \\ B^p - \frac{q}{w^p} Y^p - \frac{1}{1+\frac{1}{\beta}} \left(\frac{Y^p}{w^p} \right)^{1+\frac{1}{\beta}} &= B^{np} - \frac{1}{1+\frac{1}{\beta}} \left(\frac{Y^{np}}{w^{np}} \right)^{1+\frac{1}{\beta}}. \end{aligned}$$

where we assume that taxes are purely redistributive (i.e. there is no exogenous revenue requirement), and where the second constraint requires utility to be equalized across agents of different types.

Forming the Lagrangian of the government's problem above, and denoting by μ the Lagrange multiplier attached to the government's budget constraint and by θ the Lagrange multiplier attached to the equal-utility constraint, the first order conditions for an optimum are:

$$Y^p \quad : \quad (1-\theta) \left[\frac{q}{w^p} + \frac{1}{w^p} \left(\frac{Y^p}{w^p} \right)^{1/\beta} \right] = \mu \pi \tag{2}$$

$$B^p \quad : \quad 1-\theta = \mu \pi \tag{3}$$

$$Y^{np} \quad : \quad \theta \frac{1}{w^{np}} \left(\frac{Y^{np}}{w^{np}} \right)^{1/\beta} = (1-\pi) \mu \tag{4}$$

$$B^{np} \quad : \quad \theta = (1-\pi) \mu. \tag{5}$$

Combining (2) and (3) we get:

$$\frac{q}{w^p} + \frac{1}{w^p} \left(\frac{Y^p}{w^p} \right)^{1/\beta} = 1 \Longrightarrow Y^p = w^p \left[w^p - q \right]^\beta, \tag{6}$$

whereas combing (4) and (5) we get:

$$\frac{1}{w^{np}}\left(\frac{Y^{np}}{w^{np}}\right)^{1/\beta}=1\Longrightarrow Y^{np}=\left(w^{np}\right)^{\beta+1}. \tag{7}$$

Thus, we have that

$$Y^{p}>(<)\,Y^{np}\Longleftrightarrow w^{p}\left[w^{p}-q\right]^{\beta}>(<)\left(w^{np}\right)^{\beta+1}. \tag{8}$$

Substituting the values for Y^p and Y^{np}, given by (6) and (7), into the θ-constraint, we have:

$$B^{p}-\frac{q}{w^{p}}w^{p}\left[w^{p}-q\right]^{\beta}-\frac{1}{1+\frac{1}{\beta}}\left(w^{p}-q\right)^{\beta+1}=B^{np}-\frac{1}{1+\frac{1}{\beta}}\left(w^{np}\right)^{\beta+1},$$

and therefore:

$$B^{p}=B^{np}-\frac{\beta}{1+\beta}\left(w^{np}\right)^{\beta+1}+\frac{q+\beta w^{p}}{1+\beta}\left[w^{p}-q\right]^{\beta}. \tag{9}$$

Substituting into the μ-constraint the values for Y^p and Y^{np} given by (6) and (7), and the value for B^p given by (9) gives:

$$\pi\left[w^{p}\left(w^{p}-q\right)^{\beta}+\frac{\beta}{1+\beta}\left(w^{np}\right)^{\beta+1}-\frac{q+\beta w^{p}}{1+\beta}\left(w^{p}-q\right)^{\beta}\right]+(1-\pi)\left(w^{np}\right)^{\beta+1}=B^{np},$$

which implies:

$$\begin{aligned}B^{np} &= \pi\frac{(1+\beta)\,w^{p}\left(w^{p}-q\right)^{\beta}-\left(q+\beta w^{p}\right)\left(w^{p}-q\right)^{\beta}+\beta\left(w^{np}\right)^{\beta+1}}{1+\beta}+(1-\pi)\left(w^{np}\right)^{\beta+1},\\ &= \frac{\pi\left(w^{p}-q\right)^{\beta+1}+(1+\beta-\pi)\left(w^{np}\right)^{\beta+1}}{1+\beta}\end{aligned}$$

and therefore:

$$\begin{aligned}B^{p} &= \frac{\pi\left(w^{p}-q\right)^{\beta+1}+(1+\beta-\pi)\left(w^{np}\right)^{\beta+1}-\beta\left(w^{np}\right)^{\beta+1}+\left(q+\beta w^{p}\right)\left[w^{p}-q\right]^{\beta}}{1+\beta}\\ &= \frac{(1-\pi)\left(w^{np}\right)^{\beta+1}+\left(q+\beta w^{p}+\pi w^{p}-\pi q\right)\left(w^{p}-q\right)^{\beta}}{1+\beta}.\end{aligned}$$

When $Y^{np} - B^{np} > 0$, i.e. when

$$\frac{(1+\beta)\left(w^{np}\right)^{\beta+1} - \pi\left(w^{p}-q\right)^{\beta+1} - (1+\beta-\pi)\left(w^{np}\right)^{\beta+1}}{1+\beta} > 0$$

and therefore

$$\left(w^{np}\right)^{\beta+1} > \left(w^{p}-q\right)^{\beta+1}, \tag{10}$$

redistribution is directed towards parents. Thus, when

$$w^{p}\left[w^{p}-q\right]^{\beta} > \left(w^{np}\right)^{\beta+1} > \left(w^{p}-q\right)^{\beta+1}, \tag{11}$$

we have that $Y^{p} > Y^{np}$ and redistribution favors parents. This is implementable in a second-best setting as long as $U^{np}\left(Y^{np}, B^{np}\right) \geq U^{np}\left(Y^{p}, B^{p}\right)$, namely when:

$$\begin{aligned} & \frac{\pi\left(w^{p}-q\right)^{\beta+1} + (1+\beta-\pi)\left(w^{np}\right)^{\beta+1}}{1+\beta} - \beta\frac{1}{1+\beta}\left(w^{np}\right)^{\beta+1} \\ \geq \quad & \frac{(1-\pi)\left(w^{np}\right)^{\beta+1} + \left(q+\beta w^{p}+\pi w^{p}-\pi q\right)\left(w^{p}-q\right)^{\beta}}{1+\beta} \\ & -\beta\frac{1}{1+\beta}\left(w^{p}\right)^{1+\frac{1}{\beta}}\left(w^{p}-q\right)^{\beta+1}\left(\frac{1}{w^{np}}\right)^{1+\frac{1}{\beta}}, \end{aligned}$$

or, equivalently:

$$\begin{aligned} & \frac{\pi\left(w^{p}-q\right)^{\beta+1} + (1+\beta-\pi)\left(w^{np}\right)^{\beta+1} - \beta\left(w^{np}\right)^{\beta+1} - (1-\pi)\left(w^{np}\right)^{\beta+1}}{1+\beta} \\ & + \frac{\beta\left(w^{p}\right)^{1+\frac{1}{\beta}}\left(w^{p}-q\right)^{\beta+1}\left(\frac{1}{w^{np}}\right)^{1+\frac{1}{\beta}} - \left(q+\beta w^{p}+\pi w^{p}-\pi q\right)\left(w^{p}-q\right)^{\beta}}{1+\beta} \\ \geq \quad & 0, \end{aligned}$$

and therefore:

$$\begin{aligned} & \pi\left(w^{p}-q\right)^{\beta+1} + (1+\beta-\pi)\left(w^{np}\right)^{\beta+1} - \beta\left(w^{np}\right)^{\beta+1} - (1-\pi)\left(w^{np}\right)^{\beta+1} \\ & +\beta\left(w^{p}\right)^{1+\frac{1}{\beta}}\left(w^{p}-q\right)^{\beta+1}\left(\frac{1}{w^{np}}\right)^{1+\frac{1}{\beta}} - \left(q+\beta w^{p}+\pi w^{p}-\pi q\right)\left(w^{p}-q\right)^{\beta} \\ \geq \quad & 0, \end{aligned}$$

from which one obtains, simplifying and rearranging terms:

$$\left(w^{np}\right)^{1+\frac{1}{\beta}} \leq \frac{\beta\left(w^{p}\right)^{1+\frac{1}{\beta}}\left(w^{p}-q\right)}{q+\beta w^{p}}. \tag{12}$$

Thus, when (11) is satisfied and (12) is violated, i.e. when

$$w^{p}\left[w^{p}-q\right]^{\beta} > \left(w^{np}\right)^{\beta+1} > \max\left\{\left[\frac{\beta\left(w^{p}-q\right)}{q+\beta w^{p}}\right]^{\beta}\left(w^{p}\right)^{\beta+1},\left(w^{p}-q\right)^{\beta+1}\right\}, \tag{13}$$

the first-best max-min optimum (requiring $Y^{p} > Y^{np}$ and $Y^{p}-B^{p} < 0$) will not be implementable in a second-best setting where the government only knows the distribution of types in the population and all agents are subject to the same nonlinear income tax schedule. In this case, the following result applies:

Proposition 1 *Assume that (13) holds, so that at a first-best max-min optimum, parents earn more than non-parents, redistribution goes from low-income- to high-income earners, and the first-best max-min optimum is not implementable with an anonymous nonlinear income tax schedule. At a second-best optimum:*

i) parents will face a negative marginal tax rate, whereas the labor supply of non-parents is left undistorted;

ii) both parents and non-parents will enjoy the same level of utility when

$$\pi < \frac{q}{w^{p}}\frac{\left(w^{p}\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}}{\left[q-\beta\left(w^{p}-q\right)\right]\left(w^{p}\right)^{\frac{1}{\beta}}+\beta\left(w^{np}\right)^{1+\frac{1}{\beta}}}. \tag{14}$$

Proof. see the Appendix. ■

Before moving to the next section, where we will provide various numerical examples to illustrate the possibility that a second-best optimum features negative marginal tax rates, it is worth emphasizing that, in contrast to what happens in standard optimal tax models, the so called *single-crossing* property (also known as *agent monotonicity* property) is not satisfied in our setting. This property prescribes that, at any bundle in the (Y, B)-space, the indifference curves are flatter the higher the wage rate of an agent. In our setting, and for a given (Y, B)-bundle, parents have an indifference curve with slope equal to

$$MRS^{p}_{YB}(Y, B) = \frac{1}{w^p}\left[q - \frac{\partial u\left(B - q\frac{Y}{w^p}, \frac{Y}{w^p}\right)/\partial h}{\partial u\left(B - q\frac{Y}{w^p}, \frac{Y}{w^p}\right)/\partial c}\right] = \frac{q}{w^p} + \frac{1}{w^p}\left(\frac{Y}{w^p}\right)^{1/\beta},$$

whereas non-parents have an indifference curve with slope equal to

$$MRS^{np}_{YB}(Y, B) = -\frac{1}{w^{np}}\frac{\partial u\left(B, \frac{Y}{w^{np}}\right)/\partial h}{\partial u\left(B, \frac{Y}{w^{np}}\right)/\partial c} = \frac{1}{w^{np}}\left(\frac{Y}{w^{np}}\right)^{1/\beta}.$$

Thus, unless $w^p < w^{np}$, in which case $MRS^{p}_{YB} - MRS^{np}_{YB} > 0$ at any bundle in the (Y, B)-space, the sign of the difference $MRS^{p}_{YB} - MRS^{np}_{YB}$ will depend on the specific (Y, B)-bundle that is considered. More precisely, when $w^p > w^{np}$ we will have that $MRS^{p}_{YB} - MRS^{np}_{YB} < 0$ for values of Y such that

$$Y > \left(\frac{q}{w^p}\right)^{\beta}\left[\frac{(w^p)^{1+\frac{1}{\beta}} - (w^{np})^{1+\frac{1}{\beta}}}{(w^{np})^{1+\frac{1}{\beta}}(w^p)^{1+\frac{1}{\beta}}}\right]^{-\beta} = \left(\frac{q}{w^p}\right)^{\beta}(w^{np})^{1+\beta}(w^p)^{1+\beta}\left[(w^p)^{1+\frac{1}{\beta}} - (w^{np})^{1+\frac{1}{\beta}}\right]^{-\beta}.$$

The fact that the *single-crossing* property is not satisfied in our setting shows that our bi-dimensional heterogeneity (in skills and needs) cannot be reduced to one dimension (in contrast to what happens for instance in Choné and Laroque, 2010).

The possibility of having parents facing a negative marginal tax rate at a second-best max-min optimum is illustrated by the following numerical examples.

2.1 Numerical examples

Set $\beta = 1$, $\pi = 0.45$, $w^p = 13$, $w^{np} = 10$, $q = 5$.[5] Let the implicit marginal income tax rate faced by an agent of type j ($j = p, np$) at a bundle (Y, B) be

[5] Notice that the assumptions on w^{np}, w^p and q are consistent with the condition (13).

given by $1 - MRS^j_{YB}(Y, B)$, and let the average tax rate at income Y be given by $ATR(Y) = (Y - B)/Y$. At a first-best max-min optimum we have:

$$\begin{aligned}
Y^p &= 104.00 \quad B^p = 113.90 \quad c^p = 73.90 \quad h^p = 8.00 \\
Y^{np} &= 100.00 \quad B^{np} = 91.90 \quad c^{np} = 91.90 \quad h^{np} = 10.00 \\
T'(Y^{np}) &= T'(Y^p) = 0 \\
U^p(Y^p, B^p) &= 41.90 \quad U^{np}(Y^{np}, B^{np}) = 41.90 \quad U^{np}(Y^p, B^p) = 59.82 \\
SWF &= 41.9 \quad ATR(Y^p, B^p) = -9.52\% \quad ATR(Y^{np}, B^{np}) = 8.10\%
\end{aligned}$$

At a second-best max-min optimum we have:

$$\begin{aligned}
Y^p &= 156.74 \quad B^p = 165.59 \quad c^p = 105.31 \quad h^p = 12.06 \\
Y^{np} &= 100.00 \quad B^{np} = 92.76 \quad c^{np} = 92.76 \quad h^{np} = 10.00 \\
T'(Y^{np}) &= 0 \quad T'(Y^p) = -31.21\% \\
U^p(Y^p, B^p) &= 32.62 \quad U^{np}(Y^{np}, B^{np}) = 42.76 \quad U^{np}(Y^p, B^p) = 42.76 \\
SWF &= 32.62 \quad ATR(Y^p, B^p) = -5.65\% \quad ATR(Y^{np}, B^{np}) = 7.24\%
\end{aligned}$$

In the above example, the parameter values were such that

$$\pi \geq \frac{q}{w^p} \frac{(w^p)^{1+\frac{1}{\beta}} - (w^{np})^{1+\frac{1}{\beta}}}{[q - \beta(w^p - q)](w^p)^{\frac{1}{\beta}} + \beta(w^{np})^{1+\frac{1}{\beta}}} \simeq 43.51\%,$$

and according to the result stated in the Proposition, we find that non-parents enjoy a higher level of utility than parents. Lowering π to, e.g., 0.4 one would get the result that utilities are equalized also at a second-best optimum. In particular, with $\pi = 0.4$ we would have that at a first-best max-min optimum:

$$\begin{aligned}
Y^p &= 104.00 \quad B^p = 114.80 \quad c^p = 74.80 \quad h^p = 8.00 \\
Y^{np} &= 100.00 \quad B^{np} = 92.80 \quad c^{np} = 92.80 \quad h^{np} = 10 \\
T'(Y^{np}) &= T'(Y^p) = 0 \\
U^p(Y^p, B^p) &= 42.8 \quad U^{np}(Y^{np}, B^{np}) = 42.80 \quad U^{np}(Y^p, B^p) = 60.72 \\
SWF &= 42.80 \quad ATR(Y^p, B^p) = -10.38\% \quad ATR(Y^{np}, B^{np}) = 7.20\%
\end{aligned}$$

whereas at a second-best optimum:

$$\begin{aligned}
Y^p &= 188.41 \quad B^p = 211.85 \quad c^p = 139.39 \quad h^p = 14.49 \\
Y^{np} &= 100.00 \quad B^{np} = 84.37 \quad c^{np} = 84.37 \quad h^{np} = 10 \\
T'(Y^{np}) &= 0 \quad T'(Y^p) = -49.94\% \\
U^p(Y^p, B^p) &= 34.37 \quad U^{np}(Y^{np}, B^{np}) = 34.37 \quad U^{np}(Y^p, B^p) = 34.37 \\
SWF &= 34.37 \quad ATR(Y^p, B^p) = -12.44\% \quad ATR(Y^{np}, B^{np}) = 15.63\%
\end{aligned}$$

In the next section we consider some extensions to our baseline model to convey the idea that the possibility of optimal negative marginal tax rates is not confined to knife-edge cases.

3 Extensions

The first assumption that we relax is the one about the value for the elasticity of labor supply. In the examples above we have considered a unitary value for the elasticity of labor supply, which can be regarded as quite large. As we show below the qualitative results that we have obtained also hold when we consider lower values for the elasticity. The second assumption that we relax is the one pertaining to the number of types. Under the two-type model analyzed in the previous section, negative marginal tax rates can only arise at the top of the income distribution. With more than two types, instead, we will see that negative marginal tax rates can also occur at intermediate levels of income (even though never at the bottom of the income distribution). The third assumption that we relax is the one pertaining to the income effects on labor supply, which were ruled out by the utility function on which the analysis of the previous section was based. This will also offer us the possibility to consider social welfare functions exhibiting a lower degree of social aversion to inequality than the max-min.

3.1 Changing the elasticity of labor supply

Reconsider the first example of the previous section (where $\pi = 0.45$, $w^p = 13$, $w^{np} = 10$, and $q = 5$) but this time lower the elasticity of labor supply from $\beta = 1$ to $\beta = 1/3$. At a first-best max-min optimum we have:

$$\begin{aligned}
Y^p &= 26.00 \quad B^p = 28.49 \quad c^p = 18.49 \quad h^p = 2.00 \\
Y^{np} &= 21.54 \quad B^{np} = 19.88 \quad c^{np} = 19.88 \quad h^{np} = 2.15 \\
T'(Y^{np}) &= T'(Y^p) = 0 \\
U^p(Y^p, B^p) &= 14.49 \quad U^{np}(Y^{np}, B^{np}) = 14.49 \quad U^{np}(Y^p, B^p) = 14.49 \\
SWF &= 14.49 \quad ATR(Y^p, B^p) = -9.59\% \quad ATR(Y^{np}, B^{np}) = 7.72\%
\end{aligned}$$

At a second-best max-min optimum we instead have:

$$\begin{aligned}
Y^p &= 28.71 \quad B^p = 31.38 \quad c^p = 20.33 \quad h^p = 2.21 \\
Y^{np} &= 21.54 \quad B^{np} = 19.77 \quad c^{np} = 19.77 \quad h^{np} = 2.15 \\
T'(Y^{np}) &= 0 \quad T'(Y^p) = -21.35\% \\
U^p(Y^p, B^p) &= 14.38 \quad U^{np}(Y^{np}, B^{np}) = 14.38 \quad U^{np}(Y^p, B^p) = 14.38 \\
SWF &= 14.38 \quad ATR(Y^p, B^p) = -9.27\% \quad ATR(Y^{np}, B^{np}) = 8.24\%
\end{aligned}$$

Using (14) one can calculate the highest value for π associated with a second-best optimum where the utility of parents and non-parents are equalized. While with $\beta = 1$ this maximum value for π was equal to 43.51%, with $\beta = 1/3$ we obtain

$$\pi < \frac{q}{w^p} \frac{\left(w^p\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}}{\left[q - \beta\left(w^p - q\right)\right]\left(w^p\right)^{\frac{1}{\beta}} + \beta\left(w^{np}\right)^{1+\frac{1}{\beta}}} \simeq 84.39\%.$$

Thus, when $w^p = 13$, $w^{np} = 10$, and $q = 5$ we get that parents face a negative marginal tax rate equal to -21.35% for all values of π satisfying the inequality above.[6] For values of π larger than 84.39% we would still get that parents face a negative marginal tax rate at a second-best optimum, but this time the magnitude of the upward distortion on their labor supply would be smaller and non-parents would enjoy a higher utility than parents.[7]

3.2 More than two types

The examples above refer to a two-type setting where, if a negative marginal tax rate is optimal, it necessarily applies to the top-income earners. However, in a more general setting with more than two types of agents, negative marginal tax rates may as well apply to intermediate levels of income. This possibility is illustrated by the following examples where we consider a three-type setting.

Assume that there are three types of agents, non-parents earning a wage rate $w^{np} = 10$, parents earning a wage rate $w_1^p = 14$ and parents earning a wage rate $w_2^p = 15$ (where we use a subscript $j = 1, 2$ to distinguish between variables or parameters pertaining to the two different groups of parents). Moreover, let $\beta = 1$, $\pi = 3/4$ (with $\pi_1 = 1/2$ and $\pi_2 = 1/4$ denoting respectively the proportions of

[6] The fact that the negative marginal tax rate faced by parents does not change is a consequence of our assumption that there are no income effects on labor supply. This implies that the marginal tax rate only depends on the gross income Y and not on B.

[7] For instance, with $\pi = 0.9$ we would obtain that $T'\left(Y^p\right) = -10.60\%$, $U^p\left(Y^p, B^p\right) = 12.21$, $U^{np}\left(Y^{np}, B^{np}\right) = 13.58$, $U^{np}\left(Y^p, B^p\right) = 13.58$.

parents of type 1 and 2 in the total population), and $q = 5$. At a second-best max-min optimum we have:

$$
\begin{aligned}
Y_2^p &= 150.00 \quad B_2^p = 152.60 \quad c_2^p = 102.60 \quad h_2^p = 10.00 \\
Y_1^p &= 145.83 \quad B_1^p = 148.47 \quad c_1^p = 96.39 \quad h_1^p = 10.42 \\
Y^{np} &= 100.00 \quad B^{np} = 92.13 \quad c^{np} = 92.13 \quad h^{np} = 10 \\
T'\left(Y^{np}\right) &= 0 \quad T'\left(Y_1^p\right) = -10.12\% \quad T'\left(Y_2^p\right) = 0 \\
U^{np}\left(Y^{np}, B^{np}\right) &= 42.13 \quad U_1^p\left(Y_1^p, B_1^p\right) = 42.13 \quad U_2^p\left(Y_2^p, B_2^p\right) = 52.60 \\
ATR\left(Y^{np}, B^{np}\right) &= 7.87\% \quad ATR\left(Y_1^p, B_1^p\right) = -1.81\% \quad ATR\left(Y_2^p, B_2^p\right) = -1.73\% \\
SWF &= 42.13
\end{aligned}
$$

At the second-best optimum described above there are two self-selection constraints that are binding. One binding constraint relates parents of type 2 to parents of type 1: the former are indifferent between choosing the bundle intended for them by the government, i.e. the bundle $(Y_2^p, B_2^p) = (150.00,\ 152.60)$, and the bundle intended for the latter, i.e. the bundle $(Y_1^p, B_1^p) = (145.83,\ 148.47)$. The other binding self-selection constraint relates non-parents to parents of type 1: the former are indifferent between choosing the bundle intended for them by the government, i.e. the bundle $(Y^{np}, B^{np}) = (100.00,\ 92.13)$, and the bundle intended for the latter, i.e. the bundle $(Y_1^p, B_1^p) = (145.83,\ 148.47)$. Even though these two binding constraints call for distorting the labor supply of parents of type 1 in opposite directions (the first requires distorting downwards the labor supply of parents of type 1, whereas the other requires an upward distortion on the labor supply of parents of type 1), relaxing the second self-selection constraint proves to be more important (since it is from non-parents that the government collects the revenue used to finance the redistributive cash transfers), which explains why parents of type 1 end up facing a negative marginal tax rate.

3.3 Income effects on labor supply

So far, due to our assumption that utility is quasi-linear in consumption, we have assumed away income effects on labor supply. To allow for income effects on labor supply, assume now that individual preferences are represented by the utility function

$$U = \log(c) - \frac{1}{1+\frac{1}{\beta}} h^{1+\frac{1}{\beta}}. \tag{15}$$

We set $\pi = 0.5$, $w^p = 13$, $w^{np} = 10$, $q = 5$ and focus on the case $\beta = 1$. At a first-best max-min optimum we have:

$$\begin{aligned}
Y^p &= 12.22 \quad B^p = 13.21 \quad c^p = 8.51 \quad h^p = 0.94 \\
Y^{np} &= 10.51 \quad B^{np} = 9.51 \quad c^{np} = 9.51 \quad h^{np} = 1.05 \\
T'(Y^{np}) &= 0 \quad T'(Y^p) = 0 \\
U^p(Y^p, B^p) &= 1.70 \quad U^{np}(Y^{np}, B^{np}) = 1.70 \quad U^{np}(Y^p, B^p) = 1.70 \\
SWF &= 1.70 \quad ATR(Y^p, B^p) = -8.17\% \quad ATR(Y^{np}, B^{np}) = 9.49\%
\end{aligned}$$

At a second-best max-min optimum we instead have:

$$\begin{aligned}
Y^p &= 14.79 \quad B^p = 15.85 \quad c^p = 10.20 \quad h^p = 1.13 \\
Y^{np} &= 10.60 \quad B^{np} = 9.44 \quad c^{np} = 9.44 \quad h^{np} = 1.06 \\
T'(Y^{np}) &= 0 \quad T'(Y^p) = -27.17\% \\
U^p(Y^p, B^p) &= 1.68 \quad U^{np}(Y^{np}, B^{np}) = 1.68 \quad U^{np}(Y^p, B^p) = 1.68348 \\
SWF &= 1.68 \quad ATR(Y^p, B^p) = -7.87\% \quad ATR(Y^{np}, B^{np}) = 10.92\%
\end{aligned}$$

Utilitarian SWF With individual preferences represented by (15), we can also show the possibility that negative marginal tax rates arise under a utilitarian social welfare function ($SWF = \pi U^p(Y^p, B^p) + (1-\pi) U^{np}(Y^{np}, B^{np})$). Choosing the same parameter values as above ($\pi = 0.5$, $w^p = 13$, $w^{np} = 10$, $q = 5$), at a first-best utilitarian optimum we have:

$$\begin{aligned}
Y^p &= 11.48 \quad B^p = 13.47 \quad c^p = 9.06 \quad h^p = 0.88 \\
Y^{np} &= 11.04 \quad B^{np} = 9.06 \quad c^{np} = 9.06 \quad h^{np} = 1.10 \\
T'(Y^{np}) &= T'(Y^p) = 0 \\
U^p(Y^p, B^p) &= 1.81 \quad U^{np}(Y^{np}, B^{np}) = 1.59 \quad U^{np}(Y^p, B^p) = 1.94 \\
SWF &= 1.70 \quad ATR(Y^p, B^p) = -17.31\% \quad ATR(Y^{np}, B^{np}) = 18\%
\end{aligned}$$

At a second-best utilitarian optimum we instead have:

$$\begin{aligned}
Y^p &= 13.04 \quad B^p = 13.52 \quad c^p = 8.51 \quad h^p = 1.00 \\
Y^{np} &= 10.24 \quad B^{np} = 9.76 \quad c^{np} = 9.76 \quad h^{np} = 1.02 \\
T'(Y^{np}) &= 0 \quad T'(Y^p) = -4.12\% \\
U^p(Y^p, B^p) &= 1.63764 \quad U^{np}(Y^{np}, B^{np}) = 1.75 \quad U^{np}(Y^p, B^p) = 1.75 \\
SWF &= 1.69577 \quad ATR(Y^p, B^p) = -3.69\% \quad ATR(Y^{np}, B^{np}) = 4.7\%
\end{aligned}$$

In the examples above parents earn more than non-parents both at the first-best and the second-best optimum. The example below illustrates the possibility of cases where, despite the fact that redistribution always goes from non-parents to parents, it goes from high- to low-income earners at a first-best optimum (where $Y^p < Y^{np}$), whereas it goes in the opposite direction at a second-best optimum (where $Y^p > Y^{np}$). Let $\pi = 0.5$, $w^p = 12$, $w^{np} = 10$, $q = 5$. At a first-best utilitarian optimum we have:

$$\begin{aligned}
Y^p &= 9.73 \quad B^p = 12.69 \quad c^p = 8.63 \quad h^p = 0.810998 \\
Y^{np} &= 11.59 \quad B^{np} = 8.63 \quad c^{np} = 8.63 \quad h^{np} = 1.16 \\
T'(Y^{np}) &= T'(Y^p) = 0 \\
U^p(Y^p, B^p) &= 1.83 \quad U^{np}(Y^{np}, B^{np}) = 1.48 \quad U^{np}(Y^p, B^p) = 2.07 \\
SWF &= 1.66 \quad ATR(Y^p, B^p) = -30.36\% \quad ATR(Y^{np}, B^{np}) = 25.50\%
\end{aligned}$$

At a second-best utilitarian optimum we have:

$$\begin{aligned}
Y^p &= 12.35 \quad B^p = 12.64 \quad c^p = 7.49 \quad h^p = 1.0292 \\
Y^{np} &= 10.14 \quad B^{np} = 9.86 \quad c^{np} = 9.86 \quad h^{np} = 1.014 \\
T'(Y^{np}) &= 0 \quad T'(Y^p) = -5.90\% \\
U^p(Y^p, B^p) &= 1.48 \quad U^{np}(Y^{np}, B^{np}) = 1.77 \quad U^{np}(Y^p, B^p) = 1.77 \\
SWF &= 1.63 \quad ATR(Y^p, B^p) = -2.31\% \quad ATR(Y^{np}, B^{np}) = 2.81\%
\end{aligned}$$

The intuition for the possibility that, when moving from a first-best optimum to a second-best optimum, a re-ranking of income may occur is the following. Assume, as in the previous example, that parents earn less than non-parents at the first-best utilitarian optimum and redistribution goes from high-income to low-income earners. If the first-best optimum is not incentive-compatible, an information rent must be granted to non-parents at a second-best optimum (to ensure implementability under an anonymous nonlinear income tax schedule). This implies that, while the labor supply of non-parents remains undistorted, their utility must increase compared to the first-best optimum; on the other hand, the labor supply of parents needs to be distorted and their utility must decrease compared to the first-best optimum. Since to raise the utility of non-parents, while keeping their labor supply undistorted, one has to lower Y^{np} while at the same time raise B^{np},[8] it follows that one can conceive of cases where $Y^{np} > Y^p$ and $Y^{np} - B^{np} > 0$ at a first-best optimum whereas a second-best optimum features $Y^{np} < Y^p$ and an upward distortion on the labor supply of parents.

Finally, the last example shows that, also under a utilitarian social welfare function, negative marginal tax rates do not necessarily arise at the top of the income distribution. Assume that there are three types of agents, non-parents earning a wage rate $w^{np} = 10$, parents earning a wage rate $w_1^p = 12$ and parents earning a wage rate $w_2^p = 13.5$ (where we use a subscript $j = 1, 2$ to distinguish between variables or parameters pertaining to the two different groups of parents). Moreover, let $\pi = 0.75$ (with $\pi_1 = 0.25$ and $\pi_2 = 0.5$ denoting respectively the proportions of parents of type 1 and 2 in the total population), and $q = 5$. At a second-best utilitarian optimum we have:

$$\begin{aligned}
Y_2^p &= 13.41 \quad B_2^p = 13.52 \quad c_2^p = 8.55719 \quad h_2^p = 0.99 \\
Y_1^p &= 12.37 \quad B_1^p = 12.53 \quad c_1^p = 7.38 \quad h_1^p = 1.03 \\
Y^{np} &= 10.20 \quad B^{np} = 9.81 \quad c^{np} = 9.81 \quad h^{np} = 1.02 \\
T'(Y^{np}) &= 0 \quad T'(Y_1^p) = -5.04\% \quad T'(Y_2^p) = 0 \\
U^{np}(Y^{np}, B^{np}) &= 1.76 \quad U_1^p(Y_1^p, B_1^p) = 1.47 \quad U_2^p(Y_2^p, B_2^p) = 1.65 \\
ATR(Y^{np}, B^{np}) &= -3.82\% \quad ATR(Y_1^p, B_1^p) = -1.31\% \quad ATR(Y_2^p, B_2^p) = -0.85\% \\
SWF &= 1.63
\end{aligned}$$

At the second-best optimum described above there are two self-selection constraints that are binding. One binding self-selection constraint relates parents of type 2 to parents of type 1. In particular, parents of type 2 are indiffer-

[8] This comes from the fact that, with preferences represented by (15), the set of (Y, B)-bundles where the labor supply of non-parents is undistorted satisfy the condition $\frac{1}{(w^{np})^{(1+\beta)/\beta}} BY^{1/\beta} = 1$. Thus, starting from the bundle assigned to non-parents at a first best utilitarian optimum, raising the utility of non-parents while keeping their labor supply undistorted requires to raise B and lower Y.

ent between choosing the bundle intended for them by the government, i.e. the bundle $(Y_2^p, B_2^p) = (13.41,\ 13.52)$, and the bundle intended for parents of type 1, i.e. the bundle $(Y_1^p, B_1^p) = (12.37,\ 12.53)$. The other binding self-selection constraint relates parents of type 1 to non-parents. In particular, non-parents are indifferent between choosing the bundle intended for them by the government, i.e. the bundle $(Y^{np}, B^{np}) = (10.20,\ 9.81)$, and the bundle intended for parents of type 1, i.e. the bundle $(Y_1^p, B_1^p) = (12.37,\ 12.53)$. Even though these two binding self-selection constraints call for distorting the labor supply of parents of type 1 in different directions (the first binding self-selection constraint calls for distorting downwards the labor supply of parents of type 1, whereas the second binding self-selection constraint calls for distorting upwards the labor supply of parents of type 1), relaxing the second self-selection constraint appears more relevant, which explains why parents of type 1 are subject to a negative marginal tax rate.

4 Concluding remarks

Previous contributions in the optimal tax literature have highlighted the possibility to obtain negative marginal income tax rates when introducing heterogeneous preferences or introducing an extensive margin of labor supply together with heterogeneous fixed costs of work.

In this paper we have highlighted how negative marginal income tax rates can be generated by introducing heterogeneity in needs (for a work-related consumption good) in a pure intensive-margin optimal income tax model where agents have identical preferences.

We have shown that the result holds for both a max-min- and a utilitarian social welfare function. Moreover, while in a two-type setting the possibility of having optimal negative marginal tax rates is confined to the top of the income distribution, in a setting with more than two types negative marginal tax rates can also apply at intermediate levels of income. Finally, we have shown that, when an optimal nonlinear income tax features (at least some) negative marginal tax rates, a re-ranking of income may occur when moving from a first-best optimum to a second-best optimum.

A Proof of Proposition 1

We have already established that, under our assumption (13), the first-best max-min optimum cannot be implemented via a nonlinear income tax schedule since non-parents would be better off choosing the bundle intended for parents. Therefore, in a second-best setting, the problem solved by the government becomes:

$$\max_{Y^p, B^p, Y^{np}, B^{np}} \quad B^p - \frac{q}{w^p} Y^p - \frac{1}{1+\frac{1}{\beta}} \left(\frac{Y^p}{w^p} \right)^{1+\frac{1}{\beta}}$$

subject to:

$$\begin{aligned}
\pi \left[Y^p - B^p \right] + (1-\pi) \left[Y^{np} - B^{np} \right] &= 0, \\
B^{np} - \frac{1}{1+\frac{1}{\beta}} \left(\frac{Y^{np}}{w^{np}} \right)^{1+\frac{1}{\beta}} &= B^p - \frac{1}{1+\frac{1}{\beta}} \left(\frac{Y^p}{w^{np}} \right)^{1+\frac{1}{\beta}}, \\
B^{np} - \frac{1}{1+\frac{1}{\beta}} \left(\frac{Y^{np}}{w^{np}} \right)^{1+\frac{1}{\beta}} &\geq B^p - \frac{q}{w^p} Y^p - \frac{1}{1+\frac{1}{\beta}} \left(\frac{Y^p}{w^p} \right)^{1+\frac{1}{\beta}}
\end{aligned}$$

where the second constraint is the incentive-compatibility constraint requiring that non-parents have no incentive to choose the (Y, B)-bundle intended for parents.

Denoting by μ the Lagrange multiplier attached to the government's budget constraint, by λ the multiplier attached to the incentive-compatibility constraint, and by θ the multiplier attached to the last constraint, the first order conditions for an optimum are:

$$\begin{aligned}
Y^{np} &: \quad -(\lambda+\theta)\frac{1}{w^{np}}\left(\frac{Y^{np}}{w^{np}}\right)^{\frac{1}{\beta}} = -\mu(1-\pi) && \text{(A1)}\\
B^{np} &: \quad \lambda+\theta = \mu(1-\pi) && \text{(A2)}\\
Y^{p} &: \quad \left[-\frac{q}{w^{p}} - \frac{1}{w^{p}}\left(\frac{Y^{p}}{w^{p}}\right)^{\frac{1}{\beta}}\right](1-\theta) + \lambda\frac{1}{w^{np}}\left(\frac{Y^{p}}{w^{np}}\right)^{\frac{1}{\beta}} + \mu\pi = 0 && \text{(A3)}\\
B^{p} &: \quad 1-\mu\pi = \lambda+\theta && \text{(A4)}
\end{aligned}$$

Combining (A1) and (A2) gives:

$$Y^{np} = (w^{np})^{\beta+1}, \tag{A5}$$

which implies that the labor supply of non-parents is left undistorted.

Moreover, combining (A2) and (A4) gives:

$$\begin{aligned}
\mu &= 1, && \text{(A6)}\\
\lambda &= 1-\pi-\theta. && \text{(A7)}
\end{aligned}$$

Notice that at the solution to the government's problem above, the θ-constraint might either be binding or not. If the θ-constraint is binding, then combining the λ-constraint and the θ-constraint we obtain:

$$B^{p} - \frac{1}{1+\frac{1}{\beta}}\left(\frac{Y^{p}}{w^{np}}\right)^{1+\frac{1}{\beta}} = B^{p} - \frac{q}{w^{p}}Y^{p} - \frac{1}{1+\frac{1}{\beta}}\left(\frac{Y^{p}}{w^{p}}\right)^{1+\frac{1}{\beta}}. \tag{A8}$$

In turn, (A8) implies that either

$$Y^{p} = 0, \tag{A9}$$

or

$$\begin{aligned}
Y^{p} &= \left(\frac{q}{w^{p}}\frac{1+\beta}{\beta}\right)^{\beta}\left[\left(\frac{1}{w^{np}}\right)^{1+\frac{1}{\beta}} - \left(\frac{1}{w^{p}}\right)^{1+\frac{1}{\beta}}\right]^{-\beta}\\
&= \left(\frac{q}{w^{p}}\frac{1+\beta}{\beta}\right)^{\beta}(w^{np})^{1+\beta}(w^{p})^{1+\beta}\left[(w^{p})^{1+\frac{1}{\beta}} - (w^{np})^{1+\frac{1}{\beta}}\right]^{-\beta}. && \text{(A10)}
\end{aligned}$$

In order to establish which of the two alternatives is welfare-dominating, consider the first possibility. When $Y^p = 0$, and with Y^{np} given by (A5), the resulting values for B^p and B^{np} (satisfying the public budget constraint and the incentive-compatibility constraint for non-parents) can be found by solving the following system of equations:

$$\begin{aligned} \pi\left[Y^p - B^p\right] + (1-\pi)\left[Y^{np} - B^{np}\right] &= 0, && \text{(A11)} \\ B^{np} - \frac{1}{1+\frac{1}{\beta}}\left(\frac{Y^{np}}{w^{np}}\right)^{1+\frac{1}{\beta}} &= B^p - \frac{1}{1+\frac{1}{\beta}}\left(\frac{Y^p}{w^{np}}\right)^{1+\frac{1}{\beta}}, && \text{(A12)} \end{aligned}$$

or, equivalently:

$$\begin{aligned} -\pi B^p + (1-\pi)\left[\left(w^{np}\right)^{\beta+1} - B^{np}\right] &= 0, \\ B^{np} - \beta\frac{1}{1+\beta}\left(w^{np}\right)^{\beta+1} &= B^p, \end{aligned}$$

which imply:

$$\begin{aligned} B^p &= \frac{1-\pi}{1+\beta}\left(w^{np}\right)^{\beta+1}, \\ B^{np} &= \frac{1-\pi+\beta}{1+\beta}\left(w^{np}\right)^{\beta+1}. \end{aligned}$$

Thus, with $Y^p = 0$ the utility for parents would be given by:

$$U^p = B^p - \frac{q}{w^p}Y^p - \frac{1}{1+\frac{1}{\beta}}\left(\frac{Y^p}{w^p}\right)^{1+\frac{1}{\beta}} = \frac{1-\pi}{1+\beta}\left(w^{np}\right)^{\beta+1}. \qquad \text{(A13)}$$

Now consider the other case in which the θ-constraint is binding, namely (A10). In this case, the resulting values for B^p and B^{np} can still be found by solving (A11)-(A12), which now imply

$$\begin{aligned} &\pi\left[\left(\frac{q}{w^p}\frac{1+\beta}{\beta}\right)^{\beta}\left(w^{np}\right)^{1+\beta}\left(w^p\right)^{1+\beta}\left[\left(w^p\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-\beta} - B^p\right] \\ &+ (1-\pi)\left[\left(w^{np}\right)^{\beta+1} - B^{np}\right] \\ =\ & 0, \end{aligned}$$

$$B^{np} - \beta \frac{1}{1+\beta} \left(w^{np}\right)^{\beta+1} = B^{p} - \beta \frac{1}{1+\beta} \left(\frac{q}{w^p} \frac{1+\beta}{\beta}\right)^{\beta+1} \frac{\left(w^{np}\right)^{\beta+1} \left(w^{p}\right)^{(1+\beta)\left(1+\frac{1}{\beta}\right)}}{\left[\left(w^{p}\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{\beta+1}},$$

and therefore:

$$\begin{aligned}
B^{np} &= \left(\frac{q}{w^p} \frac{1+\beta}{\beta}\right)^{\beta} \frac{\left(w^{np}\right)^{1+\beta} \left(w^{p}\right)^{1+\beta}}{\left[\left(w^{p}\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{\beta+1}} \left[\left(1 - \frac{q}{w^p}\right) \left(w^{p}\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right] \pi \\
&\quad + \frac{1-\pi+\beta}{1+\beta} \left(w^{np}\right)^{\beta+1} \\
B^{p} &= \left(\frac{q}{w^p} \frac{1+\beta}{\beta}\right)^{\beta} \frac{\left(w^{np}\right)^{1+\beta} \left(w^{p}\right)^{1+\beta}}{\left[\left(w^{p}\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{\beta+1}} \left\{\left[\pi + \frac{q}{w^p}\left(1-\pi\right)\right] \left(w^{p}\right)^{1+\frac{1}{\beta}} - \pi \left(w^{np}\right)^{1+\frac{1}{\beta}}\right\} \\
&\quad + \frac{1}{1+\beta} \left(1-\pi\right) \left(w^{np}\right)^{\beta+1}
\end{aligned}$$

Thus, with Y^p given by (A10) the utility for parents would be given by:

$$\begin{aligned}
U^{p} &= B^{p} - \frac{q}{w^p} Y^{p} - \frac{1}{1+\frac{1}{\beta}} \left(\frac{Y^p}{w^p}\right)^{1+\frac{1}{\beta}} \\
&= \left(\frac{q}{w^p} \frac{1+\beta}{\beta}\right)^{\beta} \frac{\left(w^{np}\right)^{1+\beta} \left(w^{p}\right)^{1+\beta}}{\left[\left(w^{p}\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{\beta+1}} \left\{\left[\pi + \frac{q}{w^p}\left(1-\pi\right)\right] \left(w^{p}\right)^{1+\frac{1}{\beta}} - \pi \left(w^{np}\right)^{1+\frac{1}{\beta}}\right\} \\
&\quad + \frac{1}{1+\beta} \left(1-\pi\right) \left(w^{np}\right)^{\beta+1} - \frac{q}{w^p} \left(\frac{q}{w^p} \frac{1+\beta}{\beta}\right)^{\beta} \frac{\left(w^{np}\right)^{1+\beta} \left(w^{p}\right)^{1+\beta}}{\left[\left(w^{p}\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{\beta}} \\
&\quad - \beta \frac{1}{1+\beta} \left\{\left(\frac{q}{w^p} \frac{1+\beta}{\beta}\right)^{\beta} \frac{\left(w^{np}\right)^{1+\beta} \left(w^{p}\right)^{\beta}}{\left[\left(w^{p}\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{\beta}}\right\}^{1+\frac{1}{\beta}}
\end{aligned}$$

and therefore:

$$\begin{aligned}
U^{p} &= \left(\frac{q}{w^p} \frac{1+\beta}{\beta}\right)^{\beta} \frac{\left(w^{np}\right)^{1+\beta} \left(w^{p}\right)^{1+\beta}}{\left[\left(w^{p}\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{\beta+1}} \left[\left(1 - \frac{q}{w^p}\right) \left(w^{p}\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right] \pi \\
&\quad + \frac{1}{1+\beta} \left(1-\pi\right) \left(w^{np}\right)^{\beta+1} \qquad \text{(A14)}
\end{aligned}$$

We can therefore conclude that, when the θ-constraint is binding at a second-best optimum, parents will supply a positive amount of labor when the value for utility given by (A14) exceeds the value provided by (A13), namely when:

$$\left(\frac{q}{w^p}\frac{1+\beta}{\beta}\right)^{\beta}\frac{\left(w^{np}\right)^{1+\beta}\left(w^{p}\right)^{1+\beta}}{\left[\left(w^{p}\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{\beta+1}}\left[\left(1-\frac{q}{w^p}\right)\left(w^{p}\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]\pi>0,$$

or, equivalently:

$$\left(w^{np}\right)^{1+\frac{1}{\beta}}<\frac{w^p-q}{w^p}\left(w^{p}\right)^{1+\frac{1}{\beta}}=\left(w^p-q\right)\left(w^{p}\right)^{\frac{1}{\beta}}.$$

Since the condition above is always satisfied under our initial assumption (13),[9] we can conclude that, when the θ-constraint is binding at the solution to the second-best government's problem, the labor supply of parents will be given by:

$$\frac{Y^p}{w^p}=\left(\frac{q}{w^p}\frac{1+\beta}{\beta}\right)^{\beta}\left(w^{np}\right)^{1+\beta}\left(w^{p}\right)^{\beta}\left[\left(w^{p}\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-\beta}. \tag{A15}$$

To ascertain whether the value for labor supply provided by (A15) implies a downward or upward distortion on the labor supply of parents, we need to compare it with the undistorted value for labor supply provided by (6). In particular, (A15) implies that the labor supply of parents is distorted upwards (i.e. they face a negative marginal tax rate[10]) when the following condition holds:

$$\left(\frac{q}{w^p}\frac{1+\beta}{\beta}\right)^{\beta}\left(w^{np}\right)^{1+\beta}\left(w^{p}\right)^{\beta}\left[\left(w^{p}\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-\beta}>\left[w^p-q\right]^{\beta},$$

[9] Notice that $\frac{\left[\pi+\frac{q}{w^p}(1-\pi)\right]^{\beta}}{\left[\left(\frac{q}{w^p}+\pi\right)\left(w^p\right)^{-1/\beta}\right]^{\beta}}>\left(w^p\right)^{1-\beta}\left(w^p-q\right)^{\beta}$.

[10] We follow the common practice in the optimal tax literature and define the (implicit) marginal tax rate faced by an agent of type j ($j=p,np$) at a given (Y,B)-bundle as $1-MRS^{j}_{YB}(Y,B)$, where $MRS^{j}_{YB}(Y,B)$ denotes the agent marginal rate of substitution between pre-tax income and after-tax income for an agent of type j at the bundle (Y,B). With our quasi-linear specification of the individual utility function, for $j=p$ we have $MRS^{p}_{YB}(Y,B)=\frac{q}{w^p}+\frac{1}{w^p}\left(\frac{Y}{w^p}\right)^{1/\beta}$, and for $j=np$ we have $MRS^{np}_{YB}(Y,B)=\frac{1}{w^{np}}\left(\frac{Y}{w^{np}}\right)^{1/\beta}$.

or, equivalently:

$$\begin{aligned}
\left(w^{np}\right)^{1+\beta} &> \left[\frac{w^p-q}{w^p}\right]^{\beta}\left(\frac{q}{w^p}\frac{1+\beta}{\beta}\right)^{-\beta}\left[\left(w^p\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{\beta} \\
\left(w^{np}\right)^{\frac{1+\beta}{\beta}} &> \frac{w^p-q}{q}\frac{\beta}{1+\beta}\left[\left(w^p\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right] \\
\left(w^{np}\right)^{1+\frac{1}{\beta}} &> \frac{\beta\left(w^p-q\right)}{q+w^p\beta}\left(w^p\right)^{1+\frac{1}{\beta}} \qquad \text{(A16)}
\end{aligned}$$

Noticing that the condition above is implied by our initial assumption (13),[11] we can conclude that, when the θ-constraint is binding at the solution to the second-best government's problem, the labor supply of parents is distorted upwards and they will face a negative marginal tax rate.

The next step is now to evaluate the labor supply of parents at a second-best optimum when the θ-constraint is non-binding. Before embarking in this task, however, it is useful for later purposes to notice that, substituting (A6), (A7) and (A10) into (A3) we get:

$$\begin{aligned}
&\left\{-\frac{q}{w^p}-\frac{q}{w^p}\frac{1+\beta}{\beta}\left(w^{np}\right)^{\frac{1+\beta}{\beta}}\left[\left(w^p\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-1}\right\}(1-\theta) \\
&+(1-\pi-\theta)\frac{q}{w^p}\frac{1+\beta}{\beta}\left(w^p\right)^{\frac{1+\beta}{\beta}}\left[\left(w^p\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-1}+\pi \\
=\ & 0,
\end{aligned}$$

which can be solved to provide the following value for θ:

$$\theta=\frac{\pi-\frac{q}{w^p}+\frac{q}{w^p}\frac{1+\beta}{\beta}\left[(1-\pi)\left(w^p\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{\frac{1+\beta}{\beta}}\right]\left[\left(w^p\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-1}}{q}\beta w^p. \qquad \text{(A17)}$$

[11] In fact, notice that, by raising to the power of β both sides of (A16) gives:

$$\left(w^{np}\right)^{1+\beta}>\left[\frac{\beta\left(w^p-q\right)}{q+w^p\beta}\right]^{\beta}\left(w^p\right)^{1+\beta}.$$

Given that the denominator of the expression on the right hand side of (A17) is positive, in order to have $\theta > 0$ the following condition must be fulfilled:

$$\pi + \frac{q}{w^p}\frac{1+\beta}{\beta}\left[\left(1-\pi\right)\left(w^p\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{\frac{1+\beta}{\beta}}\right]\left[\left(w^p\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-1} > \frac{q}{w^p},$$

or, equivalently:

$$\pi\left\{1 - \frac{q}{w^p}\frac{1+\beta}{\beta}\left(w^p\right)^{1+\frac{1}{\beta}}\left[\left(w^p\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-1}\right\} > -\frac{q}{w^p}\frac{1}{\beta}. \tag{A18}$$

Noticing that, under our initial assumption (13), it follows that the expression in curly brackets in (A18) is negative,[12] we can rewrite (A18) as:

$$\pi < \frac{q}{w^p}\frac{1}{\frac{q}{w^p}\left(1+\beta\right)\left(w^p\right)^{1+\frac{1}{\beta}}\left[\left(w^p\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-1} - \beta},$$

or, equivalently:

$$\pi < \frac{q}{w^p}\frac{\left(w^p\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}}{\left[q - \beta\left(w^p - q\right)\right]\left(w^p\right)^{\frac{1}{\beta}} + \beta\left(w^{np}\right)^{1+\frac{1}{\beta}}}. \tag{A19}$$

Thus, condition (A19) provides an upper bound for π, i.e. the proportion of parents in the population, in order for the θ-constraint to be binding at the solution to the government's problem.[13] Let's now consider the labor supply of parents at a

[12] This is true since

$$1 - \frac{q}{w^p}\frac{1+\beta}{\beta}\left(w^p\right)^{1+\frac{1}{\beta}}\left[\left(w^p\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-1} < 0$$

requires

$$\left(w^{np}\right)^{1+\beta} > \left(w^p\right)^{1+\beta}\left[\frac{w^p\beta - q\left(1+\beta\right)}{w^p\beta}\right]^{\beta},$$

and we also have that

$$\frac{w^p\beta - q\left(1+\beta\right)}{w^p\beta} < \frac{\beta\left(w^p - q\right)}{q + \beta w^p}.$$

[13] When $\theta > 0$, eq. (A17) can be combined with (A7) to obtain:

$$\lambda = \frac{-q\pi - \pi\beta w^p + q\left(1+\beta\right)\pi\left(w^p\right)^{1+\frac{1}{\beta}}\left[\left(w^p\right)^{1+\frac{1}{\beta}} - \left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-1}}{q}.$$

second-best optimum when the θ-constraint is non-binding. In this case ($\theta = 0$), from (A3) we obtain (taking into account that $\lambda = 1 - \pi$ from (A7) and that $\mu = 1$) a first order condition that takes the following form:

$$-\frac{q}{w^p} - \frac{1}{w^p}\left(\frac{Y^p}{w^p}\right)^{\frac{1}{\beta}} + (1-\pi)\frac{1}{w^{np}}\left(\frac{Y^p}{w^{np}}\right)^{\frac{1}{\beta}} + \pi = 0, \tag{A20}$$

with associated second order condition

$$(1-\pi)\frac{1}{\beta}\left(\frac{Y^p}{w^{np}}\right)^{\frac{1}{\beta}-1}\left(\frac{1}{w^{np}}\right)^2 - \frac{1}{\beta}\left(\frac{Y^p}{w^p}\right)^{\frac{1}{\beta}-1}\left(\frac{1}{w^p}\right)^2 < 0,$$

which in turn requires:

$$\pi > \frac{\left(\frac{1}{w^{np}}\right)^{1+\frac{1}{\beta}} - \left(\frac{1}{w^p}\right)^{1+\frac{1}{\beta}}}{\left(\frac{1}{w^{np}}\right)^{1+\frac{1}{\beta}}} = \frac{(w^p)^{1+\frac{1}{\beta}} - (w^{np})^{1+\frac{1}{\beta}}}{(w^p)^{1+\frac{1}{\beta}}}. \tag{A21}$$

Solving (A20) for Y^p gives:

$$Y^p = \left(\frac{q - \pi w^p}{w^p}\right)^\beta \left[(1-\pi)\left(\frac{1}{w^{np}}\right)^{1+\frac{1}{\beta}} - \left(\frac{1}{w^p}\right)^{1+\frac{1}{\beta}}\right]^{-\beta}. \tag{A22}$$

To ascertain whether the value for labor supply provided by (A22) implies a downward or upward distortion on the labor supply of parents, we need to compare it with the undistorted value for labor supply provided by (6). In particular, (A22) implies that the labor supply of parents is distorted upwards (i.e. they face a negative marginal tax rate) when the following condition holds:

$$\left(\frac{q - \pi w^p}{w^p}\right)^\beta \left[(1-\pi)\left(\frac{1}{w^{np}}\right)^{1+\frac{1}{\beta}} - \left(\frac{1}{w^p}\right)^{1+\frac{1}{\beta}}\right]^{-\beta} > w^p \left[w^p - q\right]^\beta,$$

or, equivalently:

$$(q - \pi w^p)^\beta \left[\frac{(1-\pi)(w^p)^{1+\frac{1}{\beta}} - (w^{np})^{1+\frac{1}{\beta}}}{(w^{np})^{1+\frac{1}{\beta}}(w^p)^{1+\frac{1}{\beta}}}\right]^{-\beta} > (w^p)^{1+\beta}\left[w^p - q\right]^\beta,$$

and therefore:

$$(w^{np})^{\beta+1}\left(q-\pi w^{p}\right)^{\beta}\left[\left(1-\pi\right)\left(w^{p}\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{-\beta}>\left[w^{p}-q\right]^{\beta}. \tag{A23}$$

Noticing that (A21) implies $\left(1-\pi\right)\left(w^{p}\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}<0$ and also $q-\pi w^{p}<0$,[14] inequality (A23) can be rewritten as:

$$(w^{np})^{\beta+1}>\frac{\left[w^{p}-q\right]^{\beta}\left[\left(1-\pi\right)\left(w^{p}\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]^{\beta}}{\left(q-\pi w^{p}\right)^{\beta}},$$

or, equivalently:

$$(w^{np})^{1+\frac{1}{\beta}}\left(q-\pi w^{p}\right)<\left[w^{p}-q\right]\left[\left(1-\pi\right)\left(w^{p}\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right].$$

Collecting the terms depending on w^{np}, the inequality above can be re-expressed as:

$$(w^{np})^{1+\frac{1}{\beta}}\left[q-\pi w^{p}+w^{p}-q\right]<\left[w^{p}-q\right]\left(1-\pi\right)\left(w^{p}\right)^{1+\frac{1}{\beta}},$$

or, equivalently:

$$(w^{np})^{1+\frac{1}{\beta}}<\left[w^{p}-q\right]\left(w^{p}\right)^{\frac{1}{\beta}}.$$

Thus, given that the condition above is implied by our initial assumption (13), we can conclude that, when the θ-constraint is not binding at the solution to the second-best government's problem, the labor supply of parents is distorted upwards and they will face a negative marginal tax rate.

Summarizing our results, we have that under the assumption (13), a second-best max-min optimum will always feature an upward distortion on the labor supply of parents. Moreover, when the proportion of parents is sufficiently low, namely when (A19) holds, the θ-constraint will be binding so that both parents

[14] The latter inequality comes from the fact that, under our initial assumption (13), we have that $q/w^{p}<\left[\left(w^{p}\right)^{1+\frac{1}{\beta}}-\left(w^{np}\right)^{1+\frac{1}{\beta}}\right]/\left(w^{p}\right)^{1+\frac{1}{\beta}}$.

and non-parents will enjoy the same level of utility.[15] When instead (A19) does not hold, the θ-constraint will not be binding and non-parents will enjoy a higher level of utility than parents. Finally, since parents were already earning more than non-parents at the first-best optimum (where neither the labor supply of parents nor that of non-parents were distorted), parents will still earn more than non-parents at the second-best optimum (given that the labor supply of the former is distorted upwards whereas the labor supply of the latter is left undistorted).

References

[1] Choné, P. and G. Laroque (2010) Negative marginal tax rates and heterogeneity. *American Economic Review* **100**, 2532-2547.

[2] da Costa, C. E. and L. Maestri (2017) Optimal Mirrleesian taxation in non-competitive labor markets. Mimeo.

[3] Hellwig, M. (2007) A contribution to the theory of optimal utilitarian income taxation. *Journal of Public Economics* **91**, 1449-1477.

[4] Saez, E. (2002) Optimal income transfer programs: intensive versus extensive labor supply responses. *Quarterly Journal of Economics* **117**, 1039-1073.

[5] Seade, J. (1977) On the shape of optimal tax schedules. *Journal of Public Economics* **7**, 203-235.

[6] Seade, J. (1982) On the sign of the optimum marginal income tax. *Review of Economic Studies* **49**, 637-643.

[15] Of course this (common) level of utility will be lower than the utility enjoyed by agents at a first-best optimum.

[7] Stiglitz, J. E. (1982) Self-selection and Pareto efficient taxation. *Journal of Public Economics* **17**, 213–240.