

Nonlinear and piece-wise linear income taxation, and the subsidization of work-related goods

Spencer Bastani* Sören Blomquist[†] Luca Micheletto[‡]

February 5, 2019

Abstract

We investigate how the social welfare gain of subsidizing work-related goods depends on whether the underlying income tax system is linear, piece-wise linear or fully nonlinear, focusing on child-care services as a paradigmatic example of goods/services that are complements with labor supply. Our quantitative analysis employs an empirically relevant labor supply model and shows that the welfare gain of an optimally chosen subsidy is negligible when the optimal income tax is restricted to be linear but about the same as under fully nonlinear taxation when the optimal income tax is restricted to be piece-wise linear. Our findings enhance the policy-relevance of the optimal tax argument in favor of providing subsidies to work-related goods and also shed light on the relative welfare gains of employing piece-wise linear rather than fully nonlinear income taxes.

Keywords: optimal income taxation; tagging; piece-wise linear taxation; child care subsidies

JEL Classification: H21; H42

1 Introduction

According to the seminal contribution of Atkinson and Stiglitz (1976), if the income tax is allowed to be optimally nonlinear, commodity taxes are a redundant policy instru-

*Department of Economics and Statistics, Linnaeus University; Linnaeus University Center for Integration and Discrimination Studies, Uppsala Center for Fiscal Studies, Uppsala Center for Labor Studies, CESifo. E-mail: spencer.bastani@lnu.se.

[†]Uppsala Center for Fiscal Studies at the Department of Economics, Uppsala University. E-mail: soeren.blomquist@nek.uu.se.

[‡]Department of Law, University of Milan; Dondena Center for Research on Social Dynamics, Bocconi University; Uppsala Center for Fiscal Studies; CESifo. E-mail: luca.micheletto@unimi.it.

ment when preferences are separable between leisure and other goods and individuals only differ in their innate market ability (skill). If the separability condition is not satisfied and the desired direction of redistribution goes from higher- to lower-skilled agents, one should use commodity taxes and subsidies to discourage the consumption of goods/services that are substitutes with labor supply and encourage the consumption of goods/services that are complements with labor supply.¹

There is now a sizable literature emphasizing that, in Mirrleesian income tax settings, subsidizing (or publicly providing) goods or services that are consumed in conjunction with labor supply, can be welfare-enhancing. The argument is that, by subsidizing (or subjecting to a relatively more lenient tax treatment) the purchase of goods or services that are complements with labor supply, one can slacken the binding incentive constraints faced by a government designing a nonlinear income tax for redistributive purposes.² These constraints arise because the government does not directly observe an individual's earning ability (skill), and therefore cannot levy taxes or transfers that are directly conditioned on innate ability. Instead, the government pursues its redistributive goals by means of an anonymous nonlinear income tax schedule, i.e. it designs a menu of combinations of gross incomes and taxes, and hence disposable income, and lets agents choose their preferred income point. To achieve redistribution, taxes on low-income earners must be lower than on high-income earners, and conceivably negative. High-skilled agents might then find attractive to mimic low-skilled agents by lowering their labour supply and earning an income qualifying for a lower tax burden. The tax schedule must then be designed in such a way that mimicking is deterred or, in other words, that it satisfies the incentive-compatibility constraints (self-selection constraints) requiring that each agent has no incentive to choose a point other than the one intended for his/her skill type on the income tax schedule set by the government.

One question that has so far been neglected in the literature is whether the welfare-

¹The result is similar in spirit to the classic ? result, albeit in a different setting. The role played by substitutes and complements to labor supply in optimal commodity taxation was subsequently clarified by ?.

²See, e.g., Blomquist and Christiansen (1995), Boadway and Marchand (1995), Cremer and Gahvari (1997), Balestrino (2000), Pirttilä and Tuomala (2002), and Blomquist et al. (2010). More recently, Koehne and Sachs (2017) analyze the desirability of providing tax breaks for work-related goods in the context of Pareto-efficient tax structures, whereas Bastani et al. (2017), and Ho and Pavoni (2016) analyze optimal child care subsidies.

enhancing effects from subsidizing work-related goods/services hinges on the ability of the government to optimize a fully nonlinear income tax, or if similar welfare gains can be obtained in settings where the government relies on less sophisticated income tax systems like the ones that we typically observe in real economies.³ The aim of this paper is to fill this gap in the literature by focusing on child-care services as a paradigmatic example of goods/services that are complements with labor supply.⁴ Besides the fact that real-world governments do not typically tax income on a fully nonlinear scale, one reason why the question that we address is interesting is that it is in principle ambiguous whether the effectiveness of subsidies to work-related goods as a welfare-enhancing instrument is increasing or decreasing in the degree of sophistication of the underlying income tax schedule. On the one hand, as the income tax becomes more sophisticated, social welfare increases and it becomes more difficult to reap welfare gains by supplementing income taxation with other policy instruments. On the other hand, as we have just remarked, as the income tax becomes more sophisticated (flexible), it becomes easier for the government to offset, for each agent, the distortionary effects generated by subsidizing the work-related good.

We employ a fairly canonical model where a subset of agents need child care services in order to work, and evaluate the welfare gains from subsidizing this work-related good under income tax schedules that exhibit different degree of sophistication. In particular, we consider (i) a linear tax system, (ii) a two-bracket piece-wise linear income tax, (iii) a four-bracket piece-wise linear income tax, and (iv) a fully nonlinear income tax.⁵ We partition the population into 'users' (parents) and 'non-users' (non-parents) of the work-related good and, based on the assumption that the attribute (parental status) identifying 'users' is publicly observable, we allow the government to select different tax schedules for each group. This allows us to see how the presence of a need for a work-related consumption good affects optimal marginal income taxes,

³As we have already mentioned, the Mirrleesian literature provides a rationale for subjecting work-related goods/services to a relatively more lenient tax treatment (compared to other goods) based on its effects on the incentive-compatibility constraints faced by the government in the design of the nonlinear income tax. But it is worth noticing that incentive constraints are also implicitly present under less sophisticated tax schedules, as for instance under a piece-wise linear tax, albeit they take a different form.

⁴For a recent empirical assessment of the relation between working hours and demand for child care services, see ?.

⁵The linear income tax literature was initiated by Sheshinski (1972). Some of the key papers on piece-wise linear taxation are Slemrod et al. (1994), Aaberge and Colombino (2013), Apps et al. (2013), and Andrienko et al. (2016).

and how the optimal tax schedules change as a subsidy for the work-related good is introduced.⁶ To simplify the analysis, we focus on the case where for each unit of labor supply, one unit of the work-related good needs to be acquired by 'users'. In other words, we assume that parents need one hour of child care services for every hour of market work.

Our contribution is mainly quantitative. We present numerical simulations comparing the welfare gains of subsidies for work-related goods under different income tax systems. We characterize individual behavior based on an empirically relevant labor supply model and we use administrative wage data from Sweden as our source of taxpayer heterogeneity.⁷ The labor supply model adopts a quadratic utility function specification, inspired by the contributions by Stern (1986) and Tuomala (2010). This utility specification produces realistic labor supply behavior and is also computationally convenient as it admits a closed form solution for the labor supply choice subject to a linear budget segment. This is especially practical when the government is optimizing piece-wise linear tax schedules and the optimal choice of each individual needs to be calculated repeatedly, as in the nonlinear budget set procedure of ?.⁸ To compute the optimal fully nonlinear tax systems, we employ a specification with a large number of discrete types, following the simulation approach outlined in Bastani (2015), which also enables us to use exactly the same representation of the wage distribution in all our simulations.

Our results indicate that the effectiveness of subsidies to work-related goods as a welfare-enhancing instrument is increasing in the degree of sophistication of the underlying income tax schedule. While under a linear income tax, the magnitude of the welfare gains obtained by subsidizing the work-related good is negligible, the welfare gains that can be achieved by subsidizing the work-related good under a piece-wise linear income tax amount to about the same as the gains that can be achieved by subsidizing the work-related good under an optimal nonlinear income tax. This finding enhances the policy-relevance of the optimal tax argument in favor for providing sub-

⁶The simpler case where all individuals have work-related consumption requirements and the government designs a single tax schedule, is of course a special case of our analysis.

⁷It is more appropriate to focus on a model of labor supply rather than a model of taxable income in the current context, as we are interested in needs for work-related consumption goods directly related to labor supply.

⁸However, in contrast to this study, we allow for tax schedules that generate non-convex budget sets.

sidies to work-related goods. The optimal value of the subsidy rate on the work-related good is in general quite large and is also (weakly) increasing in the degree of sophistication of the underlying income tax schedule. Finally, our results also shed light on the relative welfare gains of employing piece-wise linear rather than fully nonlinear income taxes, showing that piece-wise linear taxes are able to reap a major part of the welfare gains associated with fully nonlinear income taxes.

The paper is organized as follows. In section 2 we outline the nonlinear income tax problem, which serves as our theoretical benchmark. In section 3 we present the governments' problem in the case of linear and piece-wise linear tax structures. Section 4 describes our empirical calibration as well as the linear and piece-wise linear optimal tax problems. Section 5 describes our results and, finally, section 6 concludes.

2 The nonlinear income tax problem

Consider a setting where agents differ in terms of their labor productivity (wage rates) and their need for a work-related good (child care services). Those who need child care services in order to work are for simplicity labeled "parents" and those who do not need child care services are labelled "non-parents".⁹

We let Y denote the before tax labor income, given by the product between an agent's wage rate w and labor supply h . We also make the standard assumption that the policy maker can observe Y but not w or h separately. This rules out first-best personalized lump-sum taxes and transfers but allows labor income to be taxed on a nonlinear scale. Given our focus on child care services as a primary example of a work-related good/service, and given that parental status is an individual characteristic that can reasonably be regarded as publicly observable, we also assume that parental status can be used as a tag in the optimal tax problem, i.e. parents and non-parents face two distinct nonlinear income tax schedules.

The wage rate of an agent of skill type i belonging to group $j = p, np$, where p refers

⁹The empirically most relevant counterpart to what we label as "parents" is most likely, the so called "secondary earner" in couples with children in child care age, or the lone parents (of young children) when the household is not a couple. The reason is that these are the agents whose labor supply is primarily affected by the availability of child care services, and in this sense they can be singled out as the "users" of the work-related good which we use in our illustration. Thus, albeit we use for simplicity the labels "parents" and "non-parents", one has to bear in mind that for our purposes the group of parents represents a subset of real-world parents, consisting to a large extent of mothers of young children.

to parents and np refers to non-parents, is denoted by w^{ij} . Without loss of generality we assume that agents are ordered in such a way that $w^{1,j} < w^{2,j} < \dots < w^{N,j}$, $j = p, np$. The total population size is normalized to unity and the proportion of a type ij -agent in the population is denoted by π^{ij} and is known by the government. The (exogenous) per unit resource cost of child care services (which would be the price in a competitive market) is denoted by q . Non-parents do not need child care services. For parents, on the other hand, the demand for child care services is strictly related to the hours of work. Assuming that every parent has only one child, for every hour of work parents need one hour of child care services.¹⁰ Child care services do not represent a good that enters the parents' utility function directly; for them it entails a real cost of working, a good which must be acquired in order to work. Thus, in an economy without taxes and public expenditure, the opportunity cost of leisure, which governs the agents' decisions in an undistorted optimum, is equal to $\bar{w} \equiv w - q$ and w for, respectively, parents and non-parents. All agents have identical preferences over consumption (net of expenditures on child care) c and hours of work h ; these are represented by the utility function $u(c, h)$, possessing the standard properties.

2.1 Work-related good not subsidized

Let's start with a characterization of the solution to the government's problem when the work-related good is not subsidized. The government's objective is to maximize a weighted sum of agents' utilities. Based on the link between pre-tax earnings and post-tax earnings implied by the tax schedule that applies to them, agents choose labor supply to maximize their utility. This allows to implicitly express the marginal tax rates faced by agents as $T'(Y) = 1 - MRS$, where MRS denotes the marginal rate of substitution between gross labor income and consumption. Defining by $B \equiv Y - T(Y)$ the after-tax income associated with gross labor income Y , the government's problem can be equivalently stated as the problem of selecting bundles in the (Y, B) -space subject to a set of self-selection constraints and a public budget constraint. The self-selection constraints require that each agent (weakly) prefers the bundle intended for him/her rather than behaving as a mimicker by choosing a bundle intended for some other agent.

¹⁰This assumption is made for simplicity and does not affect the qualitative results.

Given that consumption is determined for parents as $C = B - qh = B - qY/w$ and for non-parents as $C = B$, we can define the agents' indirect utility at any given point in the (Y, B) -space as $V^{i,j}(B, Y) = u(B - \mathbf{1}[j = p]qY/w^{i,j}, Y/w^{i,j})$ where $\mathbf{1}[\cdot]$ denotes an indicator function. The slope of individuals' indifference curves in the (Y, B) -space is given by the *MRS* expression:

$$MRS^{i,j}(B, Y) = -\frac{V_Y^{i,j}}{V_B^{i,j}} = \frac{1}{w^{i,j}} \left[\mathbf{1}[j = p]q - \frac{\partial u \left(B - \mathbf{1}[j = p]q \frac{Y}{w^{i,j}}, \frac{Y}{w^{i,j}} \right) / \partial h}{\partial u \left(B - \mathbf{1}[j = p]q \frac{Y}{w^{i,j}}, \frac{Y}{w^{i,j}} \right) / \partial c} \right].$$

As can be seen from the expression above, the presence of a need for the work-related good affects the shape of the parents' indifference curves. As a consequence, and in contrast to what happens in models where agents differ only in terms of skills, (weak) normality of c is no longer a sufficient condition to ensure that, at any given point in the (Y, B) -space, the indifference curves are flatter the higher the wage rate of an agent. Notice however that, albeit this agent-monotonicity property does not hold for the population as a whole, it still holds within each of the two groups. Thus, as we are assuming that the government is optimizing separate tax schedules for parents and non-parents, it is sufficient to restrict attention to constraints linking pairs of adjacent types when formalizing the government's problem.¹¹

Denote by α^{ij} the welfare weight used by the government for agents of type ij , with $\sum_{ij} \alpha^{ij} = 1$. Furthermore, assume that the chosen welfare weights imply that, for each of the two tagged groups, the government wants to redistribute from higher- to lower ability agents so that the only (potentially) binding self-selection constraints are those running downwards and linking pair of adjacent types. Then the problem solved by the government can be formally written as:

$$\max_{\{B^{i,j}, Y^{i,j}\}} \sum_{i=1}^N \sum_{j=p, np} \alpha^{ij} V^{i,j}(B^{i,j}, Y^{i,j})$$

subject to:

$$V^{i,j}(B^{i,j}, Y^{i,j}) \geq V^{i,j}(B^{i-1,j}, Y^{i-1,j}), i \in \{2, \dots, N\}, j \in \{p, np\} (\lambda^{i,j})$$

¹¹See Guesnerie and Seade (1982) for a further elaboration on the single-crossing condition in the discrete optimal income tax model.

and

$$\sum_{i=1}^N \sum_{j=p,np} \pi^{ij} \left[(Y^{i,j} - B^{i,j}) \right] \geq 0 \quad (\mu)$$

where Lagrange multipliers are within parentheses. The first set of constraints represent the self-selection (incentive-compatibility) constraints, and the second constraint is the government's budget constraint. Implicit in the formulation of the problem above is the idea that the possibility to tag agents based on parental status allows the government to solve two separate optimal income tax problems, one for parents and one for non-parents, with the possibility of accomplishing lump-sum inter-group transfers. Obviously, tagging is always welfare-improving compared to the case where a single tax schedule applies to the whole population. The welfare-enhancing potential of a tagging scheme derives from the fact that all self-selection constraints linking agents belonging to two separate tagged groups are eliminated.¹² In the above problem, this is reflected by the fact that we have written the self-selection constraints conditional on j .

As shown in Appendix A, manipulating the first order conditions of the above problem, the general expression for the marginal tax rate faced by a type i agent, $i \in \{1, \dots, N-1\}$, belonging to group $j = p, np$ is given by:

$$T' (Y^{ij}) = \frac{1}{\mu \pi^{ij}} \left[\lambda^{i+1,j} \widehat{V}_B^{i+1,j} \left(MRS^{i,j}(B^{i,j}, Y^{i,j}) - MRS^{i+1,j}(B^{i,j}, Y^{i,j}) \right) \right] \quad (1)$$

where $\widehat{V}_B^{i+1,j} \equiv \frac{d}{dB^{i,j}} V^{i+1,j}(B^{i,j}, Y^{i,j})$. Instead, for the highest skilled agent in each group, the standard no-distortion at the top result applies, i.e. for agents $(i, j) = (N, p)$ and $(i, j) = (N, np)$, $T^{ij} = 0$.

The result provided by (1) is a standard one in the optimal tax literature and we do not discuss it at length. It states that the only reason to distort agents' (labor supply) behavior is the presence of binding self-selection constraints. Moreover, given that the agent-monotonicity property holds within each of the two tagged groups, (1) implies that the labor supply of all agents, except the highest-skilled within each group, is distorted downwards ($T' (Y^{ij}) > 0$ for $i \in \{1, \dots, N-1\}$ and $j = p, np$).

¹²The term "tagging" was coined by Akerlof (1978) to describe the use of taxes that are contingent on personal characteristics. More recent contributions on tagging and taxation include Immonen et al. (1998), Boadway and Pestieau (2006), Blomquist and Micheletto (2008), Cremer et al. (2010), Bastani (2013), Bastani et al. (2013), Bastani et al. (2015) and ?.

Let's now consider how the government's problem is modified when nonlinear income taxation is supplemented by a child care subsidy.

2.2 Work-related good subsidized

As in our model child care services enter the individual decision problem of parents as a 'needs constraint' and are not subject to a separate individual choice, it is straightforward to show that the optimal child care subsidy is 100% when two separate nonlinear tax schedules apply to parents and non-parents. To provide an intuition for this result, suppose that a fully separating equilibrium with $Y^{1,p} < \dots < Y^{N,p}$ is achieved as a solution to the government's problem described in the previous subsection. To show that a Pareto-improvement can be obtained by supplementing income taxation with a child care subsidy, consider the following tax reform. Denote respectively by $(Y^{*j,p}, B^{*j,p})$ and $(Y^{*j,np}, B^{*j,np})$ the bundle offered to parents and non-parents of skill type $j = 1, \dots, N$ at the solution to the problem where $s = 0$ (i.e., the problem described in the previous subsection). Now introduce a child care subsidy at rate $s \in (0, 1]$ and, while leaving unchanged the set of bundles $(Y^{*i,np}, B^{*i,np})$ offered to non-parents, change the set of bundles for parents by offering the following packages: $(Y^{*1,p}, B^{*1,p} - sqY^{*1,p}/w^{1,p}), \dots, (Y^{*N,p}, B^{*N,p} - sqY^{*N,p}/w^{N,p})$.

Notice that, by keeping after the reform their labor supply at the original pre-reform level, the utility of all agents would be unaffected and the government's budget constraint would still be satisfied since the income tax payment of each type of parents has been increased just enough to cover the cost of the subsidy that they receive ($sqY^{*j,p}/w^{j,p}$ for $j = 1, \dots, N$). The only effects of the reform that are left to evaluate are those on the binding self-selection constraints.

Regarding this, no effects whatsoever are generated on the self-selection constraints that are relevant in the design of the nonlinear income tax faced by non-parents.¹³ Consider now the self-selection constraints requiring higher ability parents to be prevented from mimicking lower ability parents. After implementation of the proposed reform, the consumption that a parent of skill type j can get by mimicking a parent of skill type $j - 1$ is now lower (by the amount $sq [(Y^{*j-1,p}/w^{j-1,p}) - (Y^{*j-1,p}/w^{j,p})]$) than before

¹³This is due to the fact that non-parents do not demand child care services and the fact that tagging implies that non-parents can only choose income points on the tax schedule that apply to non-parents.

the reform, whereas the labor effort that he/she has to exert has not changed. We can therefore conclude that a child care subsidy is an unambiguously welfare-enhancing instrument in this case. Moreover, we can also notice that the consumption for a j -type parent behaving as a mimicker is lowered by an amount that is increasing in s , which in turn implies that the optimal subsidy rate is in this case 100%.¹⁴

Based on the discussion above, we can then proceed to analyze the case where our work-related good is fully subsidized. In such a setting, the indirect utility is given by $V^{i,j}(B, Y) = u(B, Y/w^{i,j})$ for both $j = p$ and $j = np$ as child care purchases no longer appear in the (private) budget constraints of parents.¹⁵ Instead, these expenditures enter the government's budget constraint. The problem solved by the government in the presence of the subsidy is given by:

$$\max_{\{B^{i,j}, Y^{i,j}\}} \sum_{i=1}^N \sum_{j=p, np} \alpha^{ij} V^{i,j}(B^{i,j}, Y^{i,j})$$

subject to:

$$V^{i,j}(B^{i,j}, Y^{i,j}) \geq V^{i,j}(B^{i-1,j}, Y^{i-1,j}), i \in \{2, \dots, N\}, j \in \{p, np\} (\lambda^{i,j})$$

and

$$\sum_{i=1}^N \sum_{j=p, np} \pi^{ij} [(Y^{i,j} - B^{i,j})] \geq q \sum_{i=1}^N \pi^{ip} \frac{Y^{i,p}}{w^{i,p}} (\mu)$$

where Lagrange multipliers appear within parentheses.

As shown in Appendix B, manipulating the first order conditions of the government's problem, a general expression for the marginal tax rate faced by a type i agent,

¹⁴In our model child care services enter the individual decision problem as a 'needs constraint' and are not subject to a separate individual choice. Thus, a child care subsidy does not distort how individuals allocate their disposable income across consumption goods. The only margin of choice that it distorts is the individual leisure-labor choice. However, under a fully nonlinear income tax, where marginal income tax rates can be varied independently at each income level, the government has enough flexibility to offset for all parents, through a proper adjustment in their income tax schedule, the distortionary effect generated by a variation in the subsidy rate.

¹⁵Notice that, as compared to the case considered in the previous subsection, the parents' indifference curves in the (Y, B) -space are likely to become flatter. This certainly happens when the agents' preferences are quasi-linear in consumption, in which case the parents' indifference curves flatten by the amount q/w . More generally, the parents' indifference curves flatten after the introduction of the child care subsidy provided that the income effects on labor supply are not very large.

$i \in \{1, \dots, N - 1\}$, belonging to group $j = p, np$ can be derived:

$$T'(Y^{ij}) = \frac{1}{\mu\pi^{ij}} \left[\lambda^{i+1,j} \widehat{V}_B^{i+1,j} \left(MRS^{i,j}(B^{i,j}, Y^{i,j}) - MRS^{i+1,j}(B^{i,j}, Y^{i,j}) \right) \right] + \mathbf{1}[j = p] \frac{q}{w^{i,p}} \quad (2)$$

where, again, $\widehat{V}_B^{i+1,j} \equiv \frac{d}{dB^{i,j}} V^{i+1,j}(B^{i,j}, Y^{i,j})$. For agents of type $(i, j) = (N, np)$ we still have that $T'^{ij} = 0$ whereas for agents of type $(i, j) = (N, p)$ we have $T'^{ij} = \frac{q}{w^{i,p}}$.

Comparing (1) and (2), it can thus immediately be seen that the only difference comes from the presence of a term $\frac{q}{w^{i,p}}$ in the expressions for the marginal tax rates faced by parents when income taxation is supplemented with a child care subsidy. The introduction of a subsidy is therefore likely to lead to an increase in the marginal tax rates for parents. However, the total distortions in the economy may in fact still be reduced. Intuitively, the q/w terms that enter the expressions for the marginal tax rates faced by parents do not represent distortionary terms but serve the same role as a market price in letting parents face the right incentives.¹⁶ At the same time, the subsidy serves the purpose of weakening the self-selection constraints thwarting the government in the design of the nonlinear income tax that applies to parents. For these constraints, the mimicking-detering-effect reduces the need to distort agents for self-selection purposes. It therefore allows to reduce the truly distortionary component (i.e. the λ -terms) in the formulas for the marginal tax rates.

Notice that the expressions for the marginal tax rates that apply to non-parents do not incorporate the q/w terms. This is important, since for them, these terms would represent a truly distortionary component. Notice also that the fact that the cost of child care is not mirrored in the expressions for the marginal tax rates that apply to non-parents does not mean that the additional resources needed to finance the child care subsidy are raised only from parents. It means that if also non-parents were to participate in the financing of the child care subsidy, the additional revenue extracted from them may to a large extent be collected in a non-distortionary way through an increase in inframarginal income tax rates.¹⁷

¹⁶It forces parents to internalize the resource cost of child care which they would face in a competitive market where child care services are privately purchased.

¹⁷Nonetheless, it would be wrong to expect that the actual values of the optimal marginal income tax rates faced by non-parents would not change once the government supplements income taxation with a child care subsidy. The reason why only the first order conditions for the optimal marginal taxes faced by parents change their form is that, in the government's budget constraint, the public outlays associated with the child care subsidy are only a function of the labor supply of households with children.

Having analyzed the role of subsidies to work-related goods under a fully nonlinear income tax, in the next section we describe the quantitative model that we employ to compare the welfare-enhancing power of subsidies under different assumptions regarding the flexibility of the income tax at disposal of the government. Before doing this, however, a final remark is in order. As we have pointed out, in our setting a 100% subsidy rate is optimal under fully nonlinear taxation. This result does not necessarily extend to the case of less sophisticated tax systems as linear- and piece-wise linear tax systems. The reason is that with less sophisticated income tax systems the government no longer has the required flexibility to fully offset for each agent the distortion on the leisure-labor choice generated by subsidizing the work-related good. Thus, even though incentive-compatibility constraints are implicitly present also under piece-wise linear income taxes,¹⁸ and therefore one can still regard a subsidy to work-related goods as an instrument exerting mimicking-detering effects, full subsidization is not necessarily optimal. This also implies that it is in principle ambiguous whether the effectiveness of subsidies to work-related goods as a welfare-enhancing instrument is increasing or decreasing in the degree of sophistication of the underlying income tax schedule. On one hand, as the income tax becomes more sophisticated, social welfare increases and it becomes more difficult to reap welfare gains by supplementing income taxation with other policy instruments. On the other hand, as we have just remarked, when the income tax becomes more sophisticated it becomes easier for the government to offset for each agent the distortionary effects generated by subsidizing the work-related good.

3 The linear and piece-wise linear tax problems

We now present the government maximization problem under a four-bracket piece-wise linear tax.¹⁹ As before, we assume that the population consists of $2N$ different

However, the change in the public budget constraint generated by the inclusion of the cost of the child care subsidy will affect the value of the Lagrange multiplier associated to the public budget constraint. In turn, this will change quantitatively also the value of the marginal tax rates for non-parents, even though their first order conditions do not change.

¹⁸Albeit they take a different form than under a fully nonlinear tax since individuals on the same budget segment are pooled together.

¹⁹Our focus on piece-wise linear tax systems is motivated by the fact that most real world tax systems take this form. One reason for the wide-spread use of piece-wise linear taxes might be that such taxes are relatively easy to understand for taxpayers which might be necessary for the tax system to be per-

types of agents with wage rates $w^{1,j} < w^{2,j} < \dots < w^{N,j}$ where $j \in \{np, p\}$. The total population size is normalized to one, and π^{ij} denotes the population share of a type (i, j) agent, $i = 1, \dots, n, j \in \{np, p\}$. The piece-wise linear tax function is described by four slope parameters t_1, t_2, t_3, t_4 , and three 'break-points' Z_i defined as the points on the x -axis where the slope of T changes. The demogrant is denoted by G . Formally, the tax function as a function of income Y , is defined as:

$$T(Y) = \begin{cases} -G + t_1 Y & Y \in [0, Z_1]; \\ -G + t_1 Z_1 + t_2 (Y - Z_1) & Y \in (Z_1, Z_2]; \\ -G + t_1 Z_1 + t_2 (Z_2 - Z_1) + t_3 (Y - Z_2) & Y \in (Z_2, Z_3]; \\ -G + t_1 Z_1 + t_2 (Z_2 - Z_1) + t_3 (Z_3 - Z_2) + t_4 (Y - Z_3) & Y > Z_3. \end{cases}$$

The set of parameters of the tax function is denoted by

$$\Theta = \{(t_1, t_2, t_3, t_4, Z_1, Z_2, Z_3, G) \mid t_i \in [0, 1], Z_3 > Z_2 > Z_1, Z_i, G > 0\}.$$

The government designs two piece-wise linear tax schedules. One for parents, and one for non-parents, denoted by $T(Y; \theta^p)$ and $T(Y; \theta^{np})$, respectively. The consumption for an individual belonging to group $j \in \{p, np\}$, with productivity w , choosing to earn an income of Y under a tax schedule described by the tax parameters θ , is given by:

$$C^j(Y; \theta, w) = Y - T(Y, \theta) - \mathbf{1}[j = p] \cdot (1 - s)q \frac{Y}{w}$$

where $s \in [0, 1]$ denotes the subsidy rate to child care.

Agents choose Y to maximize $U(C^j(Y), Y) \equiv u(C^j(Y), Y/w)$ leading to the indirect utility function:

$$V^j(\theta; w) = U(C^{*j}(\theta; w), Y^{*j}(\theta; w)).$$

Under a max-min social welfare function, the government solves the following problem:

$$\max_{(\theta^{np}, \theta^p) \in \Theta^2} \left[\min \left\{ V(\theta^{np}, w^{1,np}), V(\theta^p, w^{1,p}) \right\} \right], \quad (3)$$

ceived as transparent and legitimate. Empirically, most individuals locate in the interior of segments of piece-wise linear taxes, and hence when determining their marginal tax burden, face a much simpler calculation under a piece-wise linear tax as compared to under a fully nonlinear income tax.

subject to the resource constraint:

$$\sum_{i=1}^N \sum_{j=np,p} \pi^{ij} \left(Y^{*j}(\theta^j; w^{i,j}) - C^{*j}(\theta^j; w^{i,j}) \right) \geq q \sum_{i=1}^N \pi^{ip} \frac{Y^{*p}(\theta^p; w^{i,p})}{w^{i,p}}.$$

The solution to the problem above yields an optimal piece-wise linear tax system with associated optimized tax schedules $T^{*j} = T(Y; \theta^{*j})$, $j \in \{p, np\}$.²⁰ The solution also provides a value for the inter-group transfer, which will be denoted by $G^{np,p}$, and which can be calculated as $\sum_{i=1}^N \pi^{i,np} (Y^{*np}(\theta^{np}; w^{i,np}) - C^{*np}(\theta^{np}; w^{i,np}))$.²¹ We solve this problem using numerical optimization techniques. A similar procedure is used to solve numerically the government's problem under a two-bracket piece-wise linear tax.

The case where the tax system is linear can be thought as a limit case of the piece-wise linear structure that we have described above. Simply, in the linear case, each of the two separate tax schedule features a single income bracket and a single marginal tax rate.²²

4 Quantitative Model

In this paper we use wages as a proxy for skills and calibrate the wage distribution to Swedish register data using the population distribution. Our wage data consists of individuals who worked at least part-time in 2005. Parents are defined as women with at least one child in child care age (for Sweden this corresponds to ages one to six); non-parents are defined as all men (with and without children) and all women without any child in day care age.²³ According to this definition, in 2005 the fraction

²⁰Notice that this is not a concave programming problem. Although utility is continuous in Θ , if the tax schedule displays in some intervals marginal rate regressivity, the budget set is non-convex and tax revenue is not continuous. For this reason an algorithmic approach suited for non-smooth problems needs to be used in the numerical analysis.

²¹When $\sum_{i=1}^N \pi^{i,np} (Y^{*np}(\theta^{np}; w^{i,np}) - C^{*np}(\theta^{np}; w^{i,np})) > 0$, we have that the inter-group transfer runs from non-parents to parents ($G^{np,p} > 0$). When instead $G^{np,p} < 0$ the inter-group transfer implies a redistribution of resources from parents to non-parents.

²²The government's problem under a linear tax system is described in details in Appendix C where we also provide a characterization of the optimal marginal tax rates that apply to parents and non-parents, and a characterization of the optimal child care subsidy.

²³This choice is motivated by the fact that what we have in mind as an empirical counterpart to the label "parent" is, more properly, the so called "secondary earner" in couples with children in child care age, or the lone parents (of young children) when the household is not a couple. The reason is that these are the agents whose labor supply is primarily affected by the availability of child care services, and in

of parents in Sweden was slightly below 10%.²⁴ As an estimate of the hourly price for child care, we have chosen a price of 40% of the median wage for parents.

In order to capture empirically relevant behavioral elasticities and facilitate a tractable comparison with different optimum tax models, we choose the following quadratic specification of the direct utility function:²⁵

$$u(c, h) = \alpha c^2 + \beta(J - h)^2 + \gamma c(J - h) + \delta(J - h) + \epsilon c, \quad (4)$$

where $\alpha, \beta < 0$, $\gamma, \delta, \epsilon > 0$.²⁶ The annual time endowment J is set to 5840 hours. The labor supply function is:

$$h(w) = \frac{2J\beta + m\gamma + \delta - w(2m\alpha + J\gamma + \epsilon)}{2(w^2\alpha + \beta - w\gamma)},$$

where m is virtual income and w is the wage rate. Finally, the (uncompensated) elasticity of labor supply is:

$$\eta = \left(\frac{-w^2\alpha + \beta}{w^2\alpha + \beta - w\gamma} - \frac{2J\beta + m\gamma + \delta}{2J\beta + m\gamma + \delta - w(2\gamma\alpha + J\gamma + \epsilon)} \right).$$

We make the normalization $\alpha = -1$ and impose the constraint that the labor supply function evaluated at a (net) wage rate of zero is (on average) equal to zero. This pins down β . The remaining parameters that need to be chosen are γ , δ , and ϵ . We choose $\gamma = 0.07$, $\delta = 95$ and $\epsilon = 2000$, which produce empirically relevant substitution- and income effects on labor supply.

The uncompensated labor supply elasticity as a function of the (hourly) wage rate (denoted in SEK) is shown in the top panel of figure 1. Given that the distribution of wages for parents lies to the left of the wage distribution for non-parents, and that

this sense they can be singled out as the “users” of the subsidized private good on which we focus.

²⁴Data has been combined from three sources, ‘Flergenerationsregistret’, ‘Louise-databasen’ and ‘Lönestrukturstatistikerna’, covering men and women working in the public sector and in large companies but not in small companies. According to Statistics Sweden there were 2 143 775 women in the age of 25-60 in Sweden in 2005. Our data set includes 1 457 931 wages for women and 1 519 921 wages for men. Among women, 17.43% had at least one child in day care age. This represents 8.53% of the entire population.

²⁵A similar utility function is described by Stern (1986) as a good candidate for representing labor supply behavior. The quadratic specification has also been used by Tuomala (2010) and is computationally convenient as it permits a closed form solution for the labor supply choice. This is useful especially when dealing with piece-wise linear tax schedules.

²⁶To ensure concavity we require $4\alpha\beta - \gamma^2 > 0$.

parents are interpreted as women with small children, the parametrization is consistent with the empirical finding that the labor supply of women with small children is more responsive to taxation.²⁷ The income elasticities of labor supply are shown in panel b) and range between -0.05 and -0.08, consistent with the empirical literature, as it usually documents small income effects. Finally, in the bottom panel of figure 1 the labor supply function is graphed.²⁸ Compared to parameterizations used in the earlier optimal tax literature, we believe the implied behavioral elasticities depicted in the graphs do, by and large, match more closely estimates found in the contemporary empirical labor supply literature.²⁹

To obtain a revenue-based measure of the welfare gains attainable by subsidizing child care under different income tax systems, we consider an equivalent-variation-type of welfare gain measure, taking as a benchmark the solution to the government's problem under the linear income tax optimum.³⁰ We first calculate the minimum amount of extra revenue that should be injected into the government's budget, in the linear income tax optimum without child care subsidies, in order to achieve the same social welfare level as under a different tax system (piece-wise linear or fully nonlinear income tax, with or without child care subsidies). Once we have found this minimum amount of extra revenue, we divide it by the aggregate GDP at the linear income tax optimum without child care subsidies, to get a revenue-based measure of the welfare gains.

Regarding the social welfare function, we focus on the max-min, approximating this social welfare objective with the maximization of the demogrant. This is always a valid approach when the least well-off individual does not work.³¹ In the simulation exercises presented below, the government optimizes two separate income tax schedules for the groups (parents and non-parents), and can transfer resources across the groups. In the case of a max-min social welfare function, this implies that the utility of

²⁷See, e.g., the review of the literature provided by Meghir and Phillips (2010).

²⁸The labor supply function is evaluated at an (annual) non-labor income of $m = 150000$ (SEK) which is of the same order of magnitude as the demogrant arising endogenously in the optimal tax problems that we solve.

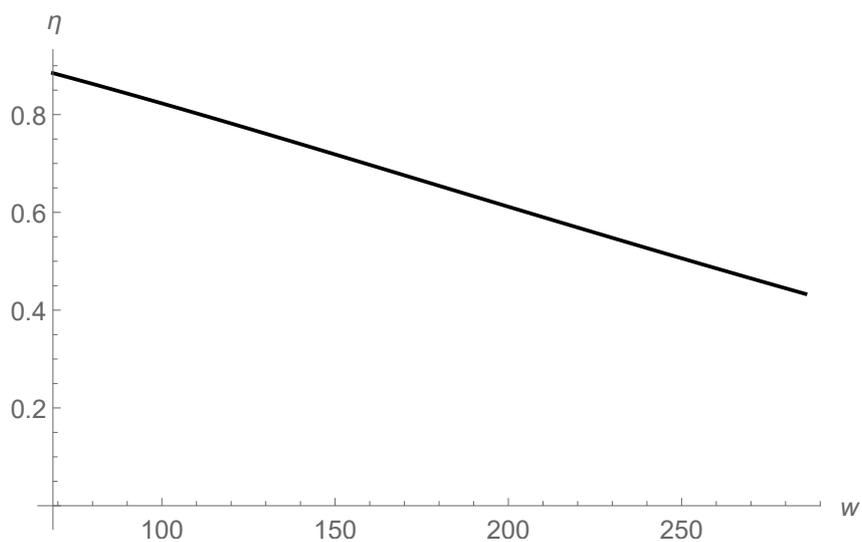
²⁹One should keep in mind that in this simulation exercise we focus on the labor supply elasticity rather than on the taxable income elasticity. It should therefore not strike as surprising that the labor supply elasticity generated by our utility function is decreasing in the wage rate of agents.

³⁰A characterization of the linear income tax optimum, with and without child care subsidies, is provided in Appendix C.

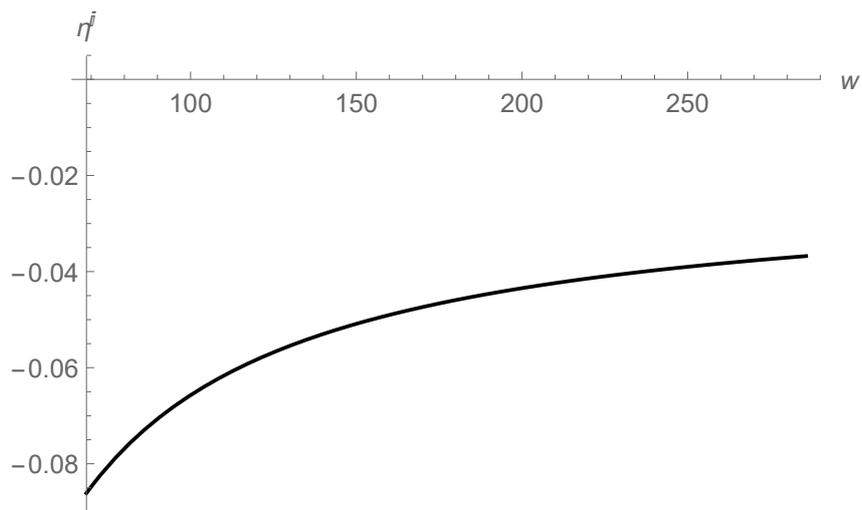
³¹In section 5.1 and appendix D we analyze the case of a (generalized) Utilitarian social welfare function.

Figure 1: Empirical targets.

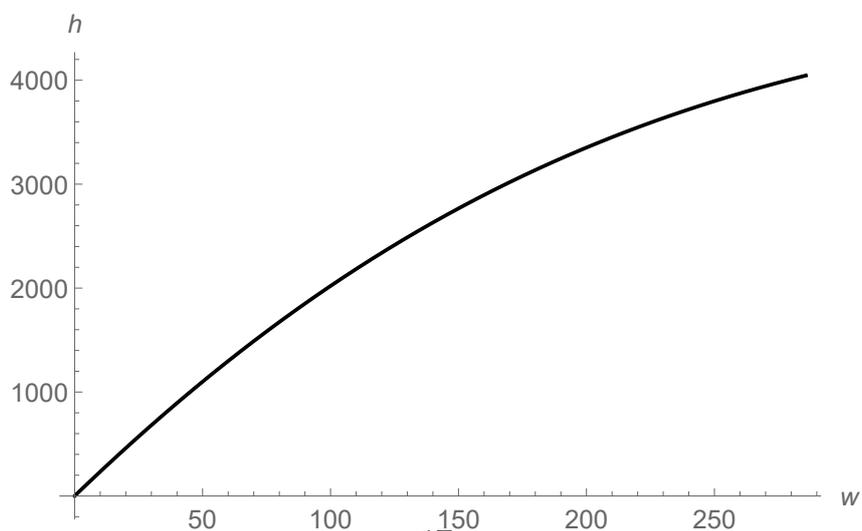
(a) The uncompensated labor supply elasticity (denoted η) over the support of the wage distribution. w =hourly wage rate (SEK).



(b) The income elasticity of labor supply (denoted η^I) over the support of the wage distribution. w =hourly wage rate (SEK).



(c) The labor supply function as a function of the hourly wage rate w (SEK).



the least well-off individual has to be the same in each group. When these agents do not work, a social welfare maximum requires the demogrant to be the same for both groups. For all the various tax systems that we consider (fully nonlinear, piece-wise linear and linear), we represent the population distribution with 1998 agents and 999 wage rates from each group. These correspond to the quantiles of each distribution, with the exclusion of the extreme values.³²

5 Quantitative Results

Our main results are contained in tables 5.1 and 5.2, and in figure 2. In the figure, we have plotted the optimal fully nonlinear income tax system together with the optimal four-bracket piece-wise linear tax system. The two top graphs display the marginal tax rate schedules for parents, whereas the two bottom graphs show the corresponding graphs for non-parents. The graphs to the left refer to the case with an optimally chosen subsidy to child care, whereas the graphs to the right refer to the case with no subsidy. The location of the break-points in the piece-wise linear tax system are indicated with vertical dashed lines. The marginal tax rates associated with the allocations chosen by agents under an optimal fully nonlinear income tax are indicated with blue dots, and the solid red line represents a kernel density approximation of the optimal schedule.³³

The values of the marginal tax rates for the linear- and piece-wise linear tax schedules are displayed in table 5.1 together with the value of the optimal subsidy rate and of the demogrant. As we can see from the table, the optimal subsidy rate drops below 100% only when the income tax system is linear. Thus, we get that in general very large subsidy rates are still optimal when the degree of sophistication of the income tax schedule is significantly lower than under a fully nonlinear income tax. At first sight this may appear counterintuitive given that, under a max-min social welfare function, the government aims at maximizing the utility of the least well-off, who are likely to

³²When approximating actual wage distributions, one can either use a set of equally-spaced wage rates together with heterogeneous probabilities, or use the percentiles of the wage distribution, along with (by construction) uniform probabilities. We have chosen the latter approach to represent the wage distribution as it makes sense to use more data points in more populated regions of the wage distribution.

³³As already mentioned, to approximate the actual wage distributions we have used the percentiles of the wage distribution, along with (by construction) uniform probabilities. This implies that some wage rates lie quite close together in regions of the wage distribution where many individuals are located. This in turn explains why we observe some bunching in the four graphs of figure 2.

be not working and therefore cannot directly benefit from a subsidy. However, notice that, since in our model all parents, irrespective of their market productivity, face an identical marginal cost of working (given by q , when $s = 0$), a proportional subsidy on work-related expenditures becomes equivalent to a progressive wage subsidy. Formally, denoting by \bar{w} the net wage rate of a parent, i.e. $\bar{w} \equiv (1 - t)w - (1 - s)q$, the combined effect of t and s is equivalent to a wage subsidy levied at rate $\bar{s} = -t + sq/w$:

$$w(1 + \bar{s}) - q = (1 - t)w - (1 - s)q \implies \bar{s} = s\frac{q}{w} - t,$$

which turns out being progressive as $\partial\bar{s}/\partial w < 0$.

Put differently, by supplementing a linear income tax with a proportional subsidy on work-related expenditures, the marginal effective income tax rate (MEITR) faced by an agent is given by $\tau \equiv t - sq/w$. Thus, even though any given individual faces a constant MEITR, the value of the MEITR is increasing in the market productivity of an agent. Despite the fact that the government does not directly observe the market productivity of an agent, the combination of a flat tax rate t and a flat subsidy s allows the government to offer parents a set of skill-dependent marginal income tax schedules. In a sense, this can be seen as the possibility to introduce in the tax system an additional layer of tagging, even though of an imperfect kind given that all parents, irrespective of their skill type, face the same demogrant, and given that the MEITR τ is constrained to vary in skill according to the function $\partial\tau/\partial w = sq/(w^2)$.

The last column of table 5.1 shows that, when supplementing the income tax that applies to parents with an optimal subsidy on work-related expenditures, there is less need to engage in inter-group redistribution (the value of $G^{np,p}$ drops in all cases when s is optimally chosen). This is due to the fact that, by relying on the subsidy the government succeeds in raising the demogrant that can be self-financed via taxation of parents' aggregate labor income.

In terms of the effects of the subsidy on the structure of optimal statutory marginal income tax rates (as opposed to MEITR), we can see from table 5.1 that, while the subsidy has minor effects on the structure of marginal tax rates for non-parents, it shifts up the structure of statutory marginal tax rates for parents. In the linear income tax case the optimal marginal tax rate increases by about 22%, from 33.29% to 55.63%. Taking into account that in our simulations q is set equal to 40% of the median wage for

parents, a subsidy at 60% coupled with an increase from 33.29% to 55.63% in t implies, roughly, that the MEITR for parents is lowered for those with a productivity below the median level and is increased for those with a productivity above the median level. For the piecewise linear income tax cases (two brackets and four brackets) we can also see that the introduction of the subsidy, rather than simply shifting up uniformly the structure of statutory marginal tax rates for parents, is accompanied by an increase in the statutory marginal tax rates that becomes smaller as one considers higher income brackets.³⁴ This implies that the statutory marginal tax rate structure faced by parents becomes more regressive as income taxation is supplemented with a subsidy on work-related expenditures.

Finally, the welfare comparisons are contained in table 5.2. The reported results show that, albeit very large subsidy rates are in general optimal, the magnitude of the welfare gains that can be achieved by using this additional policy instrument varies significantly depending on the degree of sophistication of the underlying income tax schedule. In particular, whereas the welfare gains are negligible under a linear income tax, they are roughly of the same magnitude under a piecewise linear income tax and under a fully nonlinear income tax (0.52% for the case of a two-bracket piecewise linear tax, 0.48% for the case of a four-bracket piecewise linear tax, and 0.50% for the case of a fully nonlinear income tax).

Irrespective of whether work-related expenditures are subsidized or not, table 5.2 also sheds light on the relative merits of increasingly sophisticated tax schedules. For the case when $s = 0$, the results show that, while a fully nonlinear income tax delivers large welfare gains compared to a linear income tax, a two-bracket piecewise linear tax already captures about 86% of the welfare gains achievable through a fully nonlinear optimal income tax, with the share increasing to about 91% for the case of a four-bracket piecewise linear income tax.³⁵ An almost identical picture emerges comparing the welfare gains of the various tax systems when the subsidy is optimally chosen.

³⁴For the two-bracket piecewise linear income tax case we have that $dt_1^p = 56.43\% > dt_2^p = 28.74\%$; for the two-bracket piecewise linear income tax case we have that $dt_1^p = 52.06\% > dt_2^p = 37.58\% > dt_3^p = 25.69\% > dt_4^p = 18.44\%$.

³⁵The value 86% is found dividing the welfare gains obtained under a two-bracket piecewise linear income tax (and zero subsidy), i.e. 10.91%, by the corresponding figure under a fully nonlinear income tax, i.e. 12.68%. The value 91% is obtained dividing the welfare gains under a four-bracket piecewise linear income tax (and zero subsidy), i.e. 11.62%, by the corresponding figure under a fully nonlinear income tax.

Table 5.1: Optimal linear and piece-wise linear taxes*

Linear income tax										
	s	t_1							G	$G^{np,p}$
$s = 0$										
Parents	0	33.29%							19.38	1.01
Non-parents	n/a	54.74%							19.38	0
$s \neq 0$										
Parents	0.6	55.63%							19.41	0.99
Non-parents	n/a	54.74%							19.41	0
Two bracket piece-wise linear										
	s	t_1	t_2	Z					G	$G^{np,p}$
$s = 0$										
Parents	0	34.89%	27.47%	30.99					23.03	1.31
Non-parents	n/a	89.12%	41.46%	13.11					23.03	0
$s \neq 0$										
Parents	1	91.32%	56.21%	10.74					23.20	1.16
Non-parents	n/a	90.68%	40.68%	12.96					23.20	0
Four bracket piece-wise linear										
	s	t_1	t_2	t_3	t_4	Z_1	Z_2	Z_3	G	$G^{np,p}$
$s = 0$										
Parents	0	44.13%	28.25%	32.14%	26.00%	5.69	31.92	58.00	23.26	1.31
Non-parents	n/a	95.30%	50.24%	38.74%	35.63%	9.38	26.68	146.07	23.26	0
$s \neq 0$										
Parents	1	96.19%	65.83%	57.83%	44.44%	7.02	21.27	64.00	23.42	1.15
Non-parents	n/a	94.86%	51.39%	37.69%	35.19%	9.25	26.65	161.89	23.42	0

* Notation: s denotes the subsidy rate, t_i the marginal income tax rates in bracket $i = 1, \dots, 4$, Z_i the i :th break point, $i = 1, 2, 3$. G denotes the lump-sum transfer, and $G^{np,p}$ denotes the (inter-group) transfer from non-parents to parents.

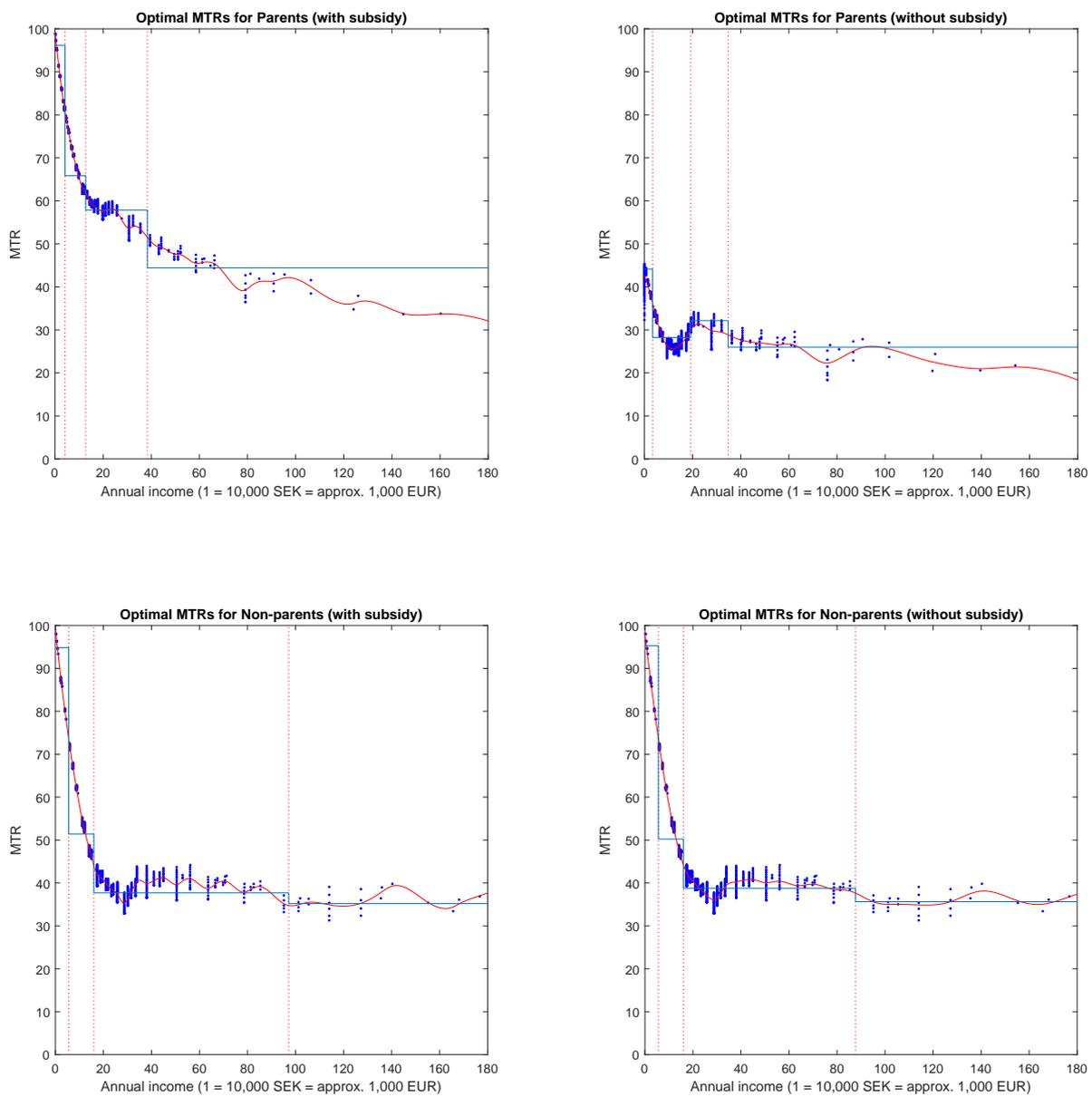


Figure 2: Optimal marginal tax rates for the fully nonlinear and four segment piecewise linear tax systems.

Table 5.2: Welfare gain comparison

Optimum	Welfare Gain
Linear ($s = 0$)	Benchmark
Linear ($s = 0.6$)	$\approx 0\%$
<i>Difference</i>	$\approx 0\%$
Two bracket piece-wise linear ($s = 0$)	10.91%
Two bracket piece-wise linear ($s = 1$)	11.43%
<i>Difference</i>	0.52%
Four bracket piece-wise linear ($s = 0$)	11.62%
Four bracket piece-wise linear ($s = 1$)	12.10%
<i>Difference</i>	0.48%
Fully Nonlinear ($s = 0$)	12.68%
Fully Nonlinear ($s = 1$)	13.18%
<i>Difference</i>	0.50%

5.1 Robustness with respect to the choice of social welfare function

Up until now we have considered a max-min social welfare objective, which in a setting where a non-zero fraction of the population is non-working, is equal to the objective of tax revenue maximization from the working population. This represents a simple and transparent benchmark case that has been analyzed extensively in the optimal tax literature. In appendix D, we examine the sensitivity of our results to the choice of social objective by examining the results when the government is maximizing the following social welfare function:

$$\sum_{i=1}^N \sum_{j=p,np} \log(V^{i,j}).$$

This is equivalent to a formulation where the social planner is of the Utilitarian type and preferences are given by $\log(u)$ where u is defined in (4).³⁶ The results from this

³⁶This is a common type of social welfare function used, for example, by Saez (2001) in his optimal tax simulations for the case when utility is quasi-linear in consumption. As is well known, the desire for redistribution in optimal tax models depends on the joint curvature of the utility function and the social welfare function. As our utility function (4) (realistically) has moderate income effects as compared to some other utility specifications in the literature, the log transformation of utility serves the purpose of introducing additional motives for redistribution.

exercise are displayed in tables D.1, D.2 and figure 3 (mirroring tables 5.1, 5.2 and figure 2). As can be seen from the summary of the welfare gains in table D.2, the gains from nonlinear income taxation under the above social welfare function specification are significantly reduced due to the substantial decrease in the governments' desire for redistribution.³⁷ However, it is still the case that the welfare gains of an optimally chosen child care subsidy is about the same under the fully nonlinear income tax as under a piece-wise linear tax.

Finally, one may also note from table D.2 that the piece-wise linear income tax is able to capture, in comparison to what happened for the case of a max-min social welfare function, a substantially smaller part (about 59% in the case of a two-bracket piece-wise linear tax, and about 66% in the case of a four-bracket piecewise linear tax) of the welfare gain associated with a fully nonlinear income tax.³⁸

6 Concluding remarks

Previous literature has shown that, in the presence of a fully nonlinear income tax, subsidizing complementary-to-labor private goods may be beneficial due to its role in alleviating the self-selection constraints faced by the government when trying to achieve redistributive goals. In this paper we have set out to examine whether this

³⁷Notice, however, that the inter-group transfer is substantially larger under the (generalized) Utilitarian social objective than under the max-min case. The reason is that in the latter case the planner was concerned with equalizing the demogrants (in order to equalize the utility of the least well off in the two groups) and therefore, except for agents at the bottom of the skill distribution, the planner did not attach weight to a reduction in the difference between the marginal utility of consumption for parents and non-parents. In the (generalized) Utilitarian case, the planner is instead concerned with the well-being of a broader set of agents, which implies that the inter-group transfer is an instrument to equalize the average net social marginal utility of income for parents and non-parents. Another difference with respect to the max-min case refers to the structure of the statutory marginal tax rates. In the absence of a subsidy, the two- and four-brackets piecewise linear taxes were generally characterized by a decreasing profile of marginal tax rates in the max-min case, a standard feature when the government's objective is the maximization of the demogrant. Under our (generalized) Utilitarian objective function, instead, the marginal tax profile is increasing in the absence of a subsidy. It becomes decreasing only for parents when an optimal subsidy is used. But also in this case the decrease in the statutory marginal tax rates over the income brackets is smaller than under the max-min objective.

³⁸The value 59% is found dividing the welfare gains under a two-bracket piecewise linear income tax (and zero subsidy), i.e. 0.26%, by the corresponding figure for the fully nonlinear income tax, i.e. 0.44%. The value 66% is found dividing the welfare gains under a four-bracket piecewise linear income tax (and zero subsidy), i.e. 0.29%, by the corresponding figure for the fully nonlinear income tax. Similar, but slightly larger, numbers would be obtained comparing piecewise linear and fully nonlinear taxes under the assumption that s is always optimally chosen. In this case a two-bracket piecewise tax would capture about 61% of the welfare gains of a fully nonlinear tax (calculated as $0.30/0.49$), and the corresponding figure for a four-bracket piecewise tax would be 69% (calculated as $0.34/0.49$).

finding is a theoretical curiosity, namely, that such gains only are achievable when the government is optimizing a fully nonlinear income tax, or, if sizable welfare gains can be obtained also when the government is optimizing simpler income tax systems of the kind used in real economies. This comparison is made possible through a computational approach where we are able to compute fully nonlinear optimal income taxes and piece-wise linear taxes under identical circumstances in terms of the components of the optimal income tax model (social welfare function, distribution of productivities, and the model of household behavior).

The message that we provide is overall positive. Using a quantitative simulation model with behavioral foundations consistent with the empirical labor supply literature, our analysis indicates that, while the effectiveness of subsidies to work-related goods as a welfare-enhancing instrument is indeed increasing in the degree of sophistication of the underlying income tax schedule, the welfare gains that can be achieved by subsidizing the work-related good under a piece-wise linear income tax is roughly the same as the gains that can be achieved by subsidizing the work-related good under an optimal fully nonlinear income tax. Regarding the optimal value of the subsidy, our results indicate that it is in general quite large and is (weakly) increasing in the degree of sophistication of the underlying income tax schedule.

Our results also indicate that, in general, an optimal nonlinear income tax delivers significant welfare gains compared to a linear income tax, even though the magnitudes of these welfare gains vary substantially depending on the chosen social welfare function. In particular, the welfare gains appear to be increasing in the degree of social aversion to inequality embedded in the social welfare function. However, in the context of our stylized model, between 66% and 91% of the gains of a fully nonlinear optimal income tax (over a linear income tax) can be captured by a four-bracket piece-wise linear income tax, depending on the choice of social welfare function, and even with a two-bracket piecewise linear tax one could capture between 59% and 86% of the gains of a fully nonlinear optimal income tax.

To conclude, we would like to emphasize that the purpose of this paper has not been to provide realistic measures of the welfare gains that can be derived from subsidizing child care in real economies, as such exercises would require a more sophisticated model of household behavior. Instead, we have used a simple and computation-

ally tractable model to illustrate how the welfare gains that derive from subsidizing a complementary-to-work good in a nonlinear income tax setting depend on the degree of sophistication of the income tax instrument.

References

- Aaberge, R. and Colombino, U. (2013). Using a Microeconometric Model of Household Labour Supply to Design Optimal Income Taxes. *Scandinavian Journal of Economics*, 115(2):449–475.
- Akerlof, G. A. (1978). The Economics of Tagging as Applied to the Optimal Income tax, Welfare Programs, and Manpower Planning. *American Economic Review*, 68:8–19.
- Andrienko, Y., Apps, P., and Rees, R. (2016). Optimal taxation and top incomes. *International Tax and Public Finance*, 23(6):981–1003.
- Apps, P., Van Long, N., and Rees, R. (2013). Optimal piecewise linear taxation. *Journal of Public Economic Theory*, 16(4):523–545.
- Atkinson, A. B. and Stiglitz, J. E. (1976). The Design of Tax Structure: Direct Versus Indirect Taxation. *Journal of Public Economics*, 6(1-2):55–75.
- Balestrino, A. (2000). Mixed tax systems and the public provision of private goods. *International Tax and Public Finance*, 7:463–478.
- Bastani, S. (2013). Gender-based and couple-based taxation. *International Tax and Public Finance*, 20(4):653–686.
- Bastani, S. (2015). Using the Discrete Model to Derive Optimal Income Tax Rates. *FinanzArchiv: Public Finance Analysis*, 71(1):106–117.
- Bastani, S., Blomquist, S., and Micheletto, L. (2013). The welfare gains of age-related optimal income taxation. *International Economic Review*, 54(4):1219–1249.
- Bastani, S., Blomquist, S., and Micheletto, L. (2017). Child Care Subsidies, Quality, and Optimal Income Taxation.

- Bastani, S., Blomquist, S., and Pirttilä, J. (2015). How should commodities be taxed? A counter-argument to the recommendation in the mirrlees review. *Oxford Economic Papers*, 67(2):455–478.
- Blomquist, S. and Christiansen, V. (1995). Public provision of private goods as a redistributive device in an optimum income tax model. *Scandinavian Journal of Economics*, 97(4):547–567.
- Blomquist, S., Christiansen, V., and Micheletto, L. (2010). Public provision of private goods and nondistortionary marginal tax rates. *American Economic Journal: Economic Policy*, 2(2):1–27.
- Blomquist, S. and Micheletto, L. (2008). Age-Related Optimal Income Taxation. *The Scandinavian Journal of Economics*, 110(2003:7):45–71.
- Boadway, R. and Marchand, M. (1995). The use of public expenditures for redistributive purposes. *Oxford Economic Papers*, 47(1):45–59.
- Boadway, R. and Pestieau, P. (2006). Tagging and Redistributive Taxation. *Annales d'Economie et de Statistique*, 83/84:123–147.
- Cremer, H. and Gahvari, F. (1997). In-kind transfers, self-selection and optimal tax policy. *European Economic Review*, 41(1):97–114.
- Cremer, H., Gahvari, F., and Lozachmeur, J. M. (2010). Tagging and Income Taxation: Theory and an Application. *American Economic Journal: Economic Policy*, 2(1):31–50.
- Guesnerie, R. and Seade, J. (1982). Nonlinear pricing in a finite economy. *Journal of Public Economics*, 17:157–179.
- Ho, C. and Pavoni, N. (2016). Efficient child care subsidies.
- Immonen, R., Kanbur, R., Keen, M., and Tuomala, M. (1998). Tagging and Taxing: The Use of Categorical and Income Information in Designing Tax/Transfer Schemes. *Economica*, 65:179–192.
- Koehne, S. and Sachs, D. (2017). Pareto Efficient Tax Breaks.

- Meghir, C. and Phillips, D. (2010). Labour supply and taxes. In *Dimensions of Tax Design: The Mirrlees Review*. OUP Oxford.
- Pirttilä, J. and Tuomala, M. (2002). Publicly provided private goods and redistribution: a general equilibrium analysis. *Scandinavian Journal of Economics*, 104:173–188.
- Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *Review of Economic Studies*, 68:205–229.
- Sheshinski, E. (1972). The optimal linear income tax. *Review of Economic Studies*, 39:297–302.
- Slemrod, J., Yitzhaki, S., Mayshar, J., and Lundholm, M. (1994). The optimal two-bracket linear income tax. *Journal of Public Economics*, 53:269–290.
- Stern, N. H. (1986). On the specification of labour supply functions. In Blundell, W. and Walker, I., editors, *Unemployment, Search and Labour Supply*, pages 121–142. Cambridge University Press, Cambridge.
- Tuomala, M. (2010). On optimal non-linear income taxation: numerical results revisited. *International Tax and Public Finance*, 17:259–270.

A Optimal marginal tax rates with optimal nonlinear income taxation

The first order conditions of the government's problem for $Y^{i,j}$ and $B^{i,j}$ (for $i \in \{1, \dots, N-1\}$ and $j = p, np$) are respectively given by.

$$\left(\alpha^{ij} + \lambda^{ij}\right) \frac{\partial V^{i,j}}{\partial Y^{i,j}} = \lambda^{i+1,j} \frac{\partial V^{i+1,j}}{\partial Y^{i,j}} - \mu \pi^{ij}, \quad (5)$$

$$\left(\alpha^{ij} + \lambda^{ij}\right) \frac{\partial V^{i,j}}{\partial B^{i,j}} = \lambda^{i+1,j} \frac{\partial V^{i+1,j}}{\partial B^{i,j}} + \mu \pi^{ij}. \quad (6)$$

Combining (5) and (6) gives:

$$\frac{\frac{\partial V^{i,j}}{\partial Y^{i,j}}}{\frac{\partial V^{i,j}}{\partial B^{i,j}}} \left[\lambda^{i+1,j} \frac{\partial V^{i+1,j}}{\partial B^{i,j}} + \mu \pi^{ij} \right] = \lambda^{i+1,j} \frac{\partial V^{i+1,j}}{\partial Y^{i,j}} - \mu \pi^{ij},$$

and therefore:

$$\mu\pi^{ij} \left[1 + \frac{\frac{\partial V^{i,j}}{\partial Y^{i,j}}}{\frac{\partial V^{i,j}}{\partial B^{i,j}}} \right] = \lambda^{i+1,j} \frac{\partial V^{i+1,j}}{\partial B^{i,j}} \left[\frac{\frac{\partial V^{i+1,j}}{\partial Y^{i,j}}}{\frac{\partial V^{i+1,j}}{\partial B^{i,j}}} - \frac{\frac{\partial V^{i,j}}{\partial Y^{i,j}}}{\frac{\partial V^{i,j}}{\partial B^{i,j}}} \right],$$

which, using the definition of (implicit) marginal tax rate $T' = 1 - MRS$, implies (1).

For $i = N$ and $j = p, np$ we instead have that the first order conditions with respect to $Y^{i,j}$ and $B^{i,j}$ are respectively given by:

$$\begin{aligned} (\alpha^{Nj} + \lambda^{N,j}) \frac{\partial V^{N,j}}{\partial Y^{N,j}} &= -\mu\pi^{Nj}, \\ (\alpha^{Nj} + \lambda^{N,j}) \frac{\partial V^{N,j}}{\partial B^{N,j}} &= \mu\pi^{Nj}, \end{aligned}$$

implying $\left[1 + \frac{\frac{\partial V^{N,j}}{\partial Y^{N,j}}}{\frac{\partial V^{N,j}}{\partial B^{N,j}}} \right] \mu\pi^{Nj} = 0$ and therefore $T'(Y^{Nij}) = 0$.

B Optimal marginal tax rates with optimal nonlinear income taxation and a child care subsidy

The first order conditions of the government's problem with respect to $Y^{i,np}$ and $B^{i,np}$ are identical to those characterizing the government's problem in the absence of a child care subsidy. Thus, the formulas characterizing the marginal tax rates faced by non-parents remain unaffected. Instead, the first order conditions of the government's problem for $Y^{i,p}$ and $B^{i,p}$ (for $i \in \{1, \dots, N-1\}$) become, respectively:

$$\left(\alpha^{ip} + \lambda^{i,p} \right) \frac{\partial V^{i,p}}{\partial Y^{i,p}} = \lambda^{i+1,p} \frac{\partial V^{i+1,p}}{\partial Y^{i,p}} - \mu\pi^{ip} + \mu q \pi^{ip} \frac{1}{w^{i,p}}, \quad (7)$$

$$\left(\alpha^{ip} + \lambda^{i,p} \right) \frac{\partial V^{i,p}}{\partial B^{i,p}} = \lambda^{i+1,p} \frac{\partial V^{i+1,p}}{\partial B^{i,p}} + \mu\pi^{ip}. \quad (8)$$

Combining (5) and (6) gives:

$$\frac{\frac{\partial V^{i,p}}{\partial Y^{i,p}}}{\frac{\partial V^{i,p}}{\partial B^{i,p}}} \left[\lambda^{i+1,p} \frac{\partial V^{i+1,p}}{\partial B^{i,p}} + \mu\pi^{ip} \right] = \lambda^{i+1,p} \frac{\partial V^{i+1,p}}{\partial Y^{i,p}} - \mu\pi^{ip} + \mu q \pi^{ip} \frac{1}{w^{i,p}},$$

and therefore:

$$\mu\pi^{ij} \left[1 + \frac{\frac{\partial V^{i,j}}{\partial Y^{i,j}}}{\frac{\partial V^{i,j}}{\partial B^{i,j}}} \right] = \lambda^{i+1,j} \frac{\partial V^{i+1,j}}{\partial B^{i,j}} \left[\frac{\frac{\partial V^{i+1,j}}{\partial Y^{i,j}}}{\frac{\partial V^{i+1,j}}{\partial B^{i,j}}} - \frac{\frac{\partial V^{i,j}}{\partial Y^{i,j}}}{\frac{\partial V^{i,j}}{\partial B^{i,j}}} \right] + \mu q \pi^{ip} \frac{1}{w^{i,p}},$$

which, using the definition of (implicit) marginal tax rate $T' = 1 - MRS$, implies (1).

For $i = N$ and $j = p$ we instead have that the first order conditions with respect to $Y^{i,p}$ and $B^{i,p}$ are respectively given by:

$$\begin{aligned} (\alpha^{Nj} + \lambda^{N,j}) \frac{\partial V^{N,j}}{\partial Y^{N,j}} &= -\mu\pi^{Nj} + \mu q \pi^{Np} \frac{1}{w^{N,p}}, \\ (\alpha^{Nj} + \lambda^{N,j}) \frac{\partial V^{N,j}}{\partial B^{N,j}} &= \mu\pi^{Nj}, \end{aligned}$$

implying $\left[1 + \frac{\partial V^{N,p}}{\partial Y^{N,p}} / \frac{\partial V^{N,p}}{\partial B^{N,p}} \right] \mu\pi^{Np} = \mu q \pi^{Np} / w^{N,p}$ and therefore $T' (Y^{Nij}) = q / w^{N,p}$.

C Characterization of an optimum under a linear tax

Under a linear income tax characterized by a marginal tax rate t^p and demogrant G^p , and supplemented by a child care subsidy levied at rate s , parents solve the problem $\max_h u(G^p + (1 - t^p)wh - (1 - s)qh, h)$. Under a linear income tax characterized by a marginal tax rate t^{np} and demogrant G^{np} , non-parents solve the problem $\max_h u(G^{np} + (1 - t^{np})wh, h)$.

Denoting respectively by $V^{i,p}(t^p, G^p, s)$ and $V^{i,np}(t^{np}, G^{np})$ the indirect utility of parents and non-parents of ability type i , and denoting by α^{ij} the welfare weight used by the government for agents of type ij , the design problem solved by the government can be written as:

$$\max_{t^p, G^p, t^{np}, G^{np}} \sum_{i=1}^N \alpha^{i,p} V^{i,p}(t^p, G^p, s) + \sum_{i=1}^N \alpha^{i,np} V^{i,np}(t^{np}, G^{np})$$

subject to:

$$\sum_{j=p,np} t^j \sum_{i=1}^N \pi^{ij} Y^{i,j} \geq \sum_{j=p,np} \sum_{i=1}^N \pi^{ij} G^j + sq \sum_{i=1}^N \pi^{ip} \frac{Y^{i,p}}{w^{i,p}}, \quad (\mu)$$

where μ is the Lagrange multiplier associated with the government's budget con-

straint.

Denote $\pi^{i,p} / \sum_{k=1}^N \pi^{k,p}$ by $\bar{\pi}^{i,p}$ and $\pi^{i,np} / \sum_{k=1}^N \pi^{k,np}$ by $\bar{\pi}^{i,np}$. The first order condition with respect to G^p and G^{np} are, respectively, given by:

$$\frac{1}{\sum_{k=1}^N \pi^{k,p}} \sum_{i=1}^N \frac{\alpha^{i,p}}{\mu} \frac{\partial V^{i,p}(t^p, G^p, s)}{\partial G^p} + \sum_{i=1}^N (t^p w^{i,p} - sq) \bar{\pi}^{i,p} \frac{\partial h^{i,p}}{\partial G^p} = 1, \quad (9)$$

$$\frac{1}{\sum_{k=1}^N \pi^{k,np}} \sum_{i=1}^N \frac{\alpha^{i,np}}{\mu} \frac{\partial V^{i,np}(t^{np}, G^{np})}{\partial G^{np}} + \sum_{i=1}^N t^{np} \bar{\pi}^{i,np} w^{i,np} \frac{\partial h^{i,np}}{\partial G^{np}} = 1. \quad (10)$$

Define the net social marginal valuation of a lump-sum transfer to a parent of type i and to a non-parent of type i as, respectively:

$$b^{i,p} \equiv \frac{1}{\sum_{k=1}^N \pi^{k,p}} \left[\frac{1}{\bar{\pi}^{i,p}} \frac{\alpha^{i,p}}{\mu} \frac{\partial V^{i,p}(t^p, G^p, s)}{\partial G^p} + (t^p w^{i,p} - sq) \frac{\partial h^{i,p}}{\partial G^p} \right],$$

$$b^{i,np} \equiv \frac{1}{\sum_{k=1}^N \pi^{k,np}} \left[\frac{1}{\bar{\pi}^{i,np}} \frac{\alpha^{i,np}}{\mu} \frac{\partial V^{i,np}(t^{np}, G^{np})}{\partial G^{np}} + (t^{np} w^{i,np}) \frac{\partial h^{i,np}}{\partial G^{np}} \right].$$

Having defined $b^{i,p}$ and $b^{i,np}$ we can easily see that condition (9)-(10) boil down to requiring $E(b^p) = E(b^{np}) = 1$, where $E(\cdot)$ denotes the expectation operator.³⁹ In other words, it prescribes that at an optimum the lump-sum component should be adjusted such that b^j , the government's net social marginal valuation of a transfer of 1 currency unit (measured in terms of government's revenue) to agents of group j (with $j = p, np$) should on average be equal to its marginal cost.

The first order condition with respect to t^{np} is the following:

$$\frac{1}{\sum_{k=1}^N \pi^{k,np}} \sum_{i=1}^N \frac{\alpha^{i,np}}{\mu} \frac{\partial V^{i,np}(t^{np}, G^{np})}{\partial t^{np}} + \mu \left[\sum_{i=1}^N \bar{\pi}^{i,np} w^{i,np} h^{i,np} + \sum_{i=1}^N t^{np} \bar{\pi}^{i,np} w^{i,np} \frac{\partial h^{i,np}}{\partial t^{np}} \right] = 0,$$

or, equivalently, applying the Slutsky equation and denoting by a tilde symbol a com-

³⁹Apart from the fact that our definition of b^i also incorporates a term depending on q , the condition $E(b) = 1$, which implicitly defines the optimal level of the demogrant, is the same that one obtains in a standard model of optimal linear income taxation without public provision.

pensated variable:

$$\frac{1}{\sum_{k=1}^N \pi^{k,np}} \sum_{i=1}^N \alpha^{i,np} \frac{\partial V^{i,np}(t^{np}, G^{np})}{\partial t^{np}} + \mu \left[\sum_{i=1}^N \bar{\pi}^{i,np} w^{i,np} h^{i,np} + \sum_{i=1}^N t^{np} \bar{\pi}^{i,np} w^{i,np} \left(\frac{\partial \tilde{h}^{i,np}}{\partial t^{np}} - w^{i,np} h^{i,np} \frac{\partial h^{i,np}}{\partial G^{np}} \right) \right] = 0.$$

Noticing that $\frac{\partial V^{i,np}(t^{np}, G^{np})}{\partial t^{np}} = -\frac{\partial V^{i,np}(t^{np}, G^{np})}{\partial G^{np}} w^{i,np} h^{i,np}$ (by applying Roy's identity) and $\frac{\partial \tilde{h}^{i,np}}{\partial t^{np}} = -w^{i,np} \frac{\partial \tilde{h}^{i,np}}{\partial w^{i,np}(1-t^{np})}$, and using (10), we can derive the following implicit expression for the optimal t^{np} :

$$\frac{t^{np}}{1-t^{np}} = -\frac{\text{cov}(b^{np}, Y^{np})}{\sum_{i=1}^N \bar{\pi}^{i,np} \gamma^{i,np} \frac{w^{i,np}(1-t^{np})}{h^{i,np}} \frac{\partial \tilde{h}^{i,np}}{\partial w^{i,np}(1-t^{np})}},$$

or, equivalently, denoting by $\tilde{\eta}_{h^{i,np}, w^{i,np}(1-t^{np})}$ the compensated elasticity of labor supply with respect to the net wage rate for a non-parent of skill type i :

$$\frac{t^{np}}{1-t^{np}} = -\frac{\text{cov}(b^{np}, Y^{np})}{\sum_{i=1}^N \bar{\pi}^{i,np} \gamma^{i,np} \tilde{\eta}_{h^{i,np}, w^{i,np}(1-t^{np})}}.$$

The first order conditions with respect to t^p and s are respectively given by:

$$\frac{1}{\sum_{k=1}^N \pi^{k,p}} \sum_{i=1}^N \alpha^{i,p} \frac{\partial V^{i,p}(t^p, G^p, s)}{\partial t^p} + \mu \left[\sum_{i=1}^N \bar{\pi}^{i,p} w^{i,p} h^{i,p} + \sum_{i=1}^N (t^p w^{i,p} - sq) \bar{\pi}^{i,p} \frac{\partial h^{i,p}}{\partial t^p} \right] = 0, \quad (11)$$

$$\frac{1}{\sum_{k=1}^N \pi^{k,p}} \sum_{i=1}^N \alpha^{i,p} \frac{\partial V^{i,p}(t^p, G^p, s)}{\partial s} + \mu \left[-q \sum_{i=1}^N \bar{\pi}^{i,p} h^{i,p} + \sum_{i=1}^N (t^p w^{i,p} - sq) \bar{\pi}^{i,p} \frac{\partial h^{i,p}}{\partial s} \right] = 0. \quad (12)$$

Using the Slutsky equation and denoting by a tilde symbol a compensated variable,

we can rewrite eqs. (11)-(12) respectively as:

$$\frac{1}{\sum_{k=1}^N \pi^{k,p}} \sum_{i=1}^N \alpha^{i,p} \frac{\partial V^{i,p}(t^p, G^p, s)}{\partial t^p} + \mu \left[\sum_{i=1}^N \bar{\pi}^{i,p} w^{i,p} h^{i,p} + \sum_{i=1}^N (t^p w^{i,p} - sq) \bar{\pi}^{i,p} \left(\frac{\partial \tilde{h}^{i,p}}{\partial t^p} - w^{i,p} h^{i,p} \frac{\partial h^{i,p}}{\partial G^p} \right) \right] = 0, \quad (13)$$

$$\frac{1}{\sum_{k=1}^N \pi^{k,p}} \sum_{i=1}^N \alpha^{i,p} \frac{\partial V^{i,p}(t^p, G^p, s)}{\partial s} + \mu \left[-q \sum_{i=1}^N \bar{\pi}^{i,p} h^{i,p} + \sum_{i=1}^N (t^p w^{i,p} - sq) \bar{\pi}^{i,p} \left(\frac{\partial \tilde{h}^{i,p}}{\partial s} + q h^{i,p} \frac{\partial h^{i,p}}{\partial G^p} \right) \right] = 0. \quad (14)$$

Notice that, by applying Roy's identity we can write $\frac{\partial V^{i,p}(t^p, G^p, s)}{\partial t^p} = -\frac{\partial V^{i,p}(t^p, G^p, s)}{\partial G^p} w^{i,p} h^{i,p}$ and $\frac{\partial V^{i,p}(t^p, G^p, s)}{\partial s} = \frac{\partial V^{i,p}(t^p, G^p, s)}{\partial G^p} q h^{i,p}$. Moreover, we have that $\frac{\partial \tilde{h}^{i,p}}{\partial t^p} = -w^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]}$ and $\frac{\partial \tilde{h}^{i,p}}{\partial s} = q \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]}$. Thus, using (9), we can rewrite (13)-(14) in matrix form as:

$$\begin{bmatrix} \sum_{i=1}^N \bar{\pi}^{i,p} (w^{i,p})^2 \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} & -\sum_{i=1}^N \bar{\pi}^{i,p} w^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} \\ \sum_{i=1}^N \bar{\pi}^{i,p} w^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} & -\sum_{i=1}^N \bar{\pi}^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} \end{bmatrix} \begin{bmatrix} t^p \\ qs \end{bmatrix} = \begin{bmatrix} -cov(b^p, Y^p) \\ -cov(b^p, h^p) \end{bmatrix} \quad (15)$$

Denoting by Δ the determinant of the 2x2 matrix on the left hand side of (15), we have:

$$\begin{aligned} \Delta &= - \left(\sum_{i=1}^N \bar{\pi}^{i,p} (w^{i,p})^2 \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} \right) \left(\sum_{i=1}^N \bar{\pi}^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} \right) \\ &\quad + \left(\sum_{i=1}^N \bar{\pi}^{i,p} w^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} \right) \left(\sum_{i=1}^N \bar{\pi}^{i,p} w^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} \right) \\ &= - \sum_{i=1}^{N-1} \left[\bar{\pi}^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} \left(\sum_{j>i} (w^{i,p} - w^{j,p})^2 \bar{\pi}^{j,p} \frac{\partial \tilde{h}^{j,p}}{\partial [w^{j,p}(1-t^p)-(1-s)q]} \right) \right] \\ &< 0. \end{aligned}$$

Thus, we have:

$$t^p = \frac{\text{cov}(b^p, Y^p) \sum_{i=1}^N \bar{\pi}^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} - \text{cov}(b^p, h^p) \sum_{i=1}^N \bar{\pi}^{i,p} w^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]}}{\Delta},$$

$$s = \frac{\text{cov}(b^p, Y^p) \sum_{i=1}^N \bar{\pi}^{i,p} w^{i,p} \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]} - \text{cov}(b^p, h^p) \sum_{i=1}^N \bar{\pi}^{i,p} (w^{i,p})^2 \frac{\partial \tilde{h}^{i,p}}{\partial [w^{i,p}(1-t^p)-(1-s)q]}}{q\Delta}.$$

D Robustness with respect to the social welfare function

Table D.1: Optimal linear and piece-wise linear taxes* (Utilitarian SWF with log)

Linear income tax										
	s	t_1				G	$G^{np,p}$			
$s = 0$										
Parents	0	5.50%				30.36	2.28			
Non-parents	n/a	12.47%				5.47	0			
$s \neq 0$										
Parents	0.765	32.80%				29.58	2.23			
Non-parents	n/a	12.47%				5.52	0			
Two bracket piece-wise linear										
	s	t_1	t_2	Z		G	$G^{np,p}$			
$s = 0$										
Parents	0	2.43%	10.22%	42.54		29.83	2.28			
Non-parents	n/a	6.33%	21.05%	77.02		3.03	0			
$s \neq 0$										
Parents	1	45.55%	31.68%	56.32		30.42	2.23			
Non-parents	n/a	6.33%	21.04%	77.02		3.09	0			
Four bracket piece-wise linear										
	s	t_1	t_2	t_3	t_4	Z_1	Z_2	Z_3	G	$G^{np,p}$
$s = 0$										
Parents	0	1.91%	4.73%	8.23%	11.34%	32.87	45.06	63.75	29.73	2.28
Non-parents	n/a	5.33%	11.71%	18.46%	23.35%	63.35	82.50	125.00	2.62	0
$s \neq 0$										
Parents	1	54.17%	44.96%	36.40%	28.58%	20.00	38.40	108.00	31.93	2.23
Non-parents	n/a	4.63%	10.13%	18.18%	23.47%	55.72	81.12	125.31	2.42	0

* Notation: s denotes the subsidy rate, t_i the marginal income tax rates in bracket $i = 1, \dots, 4$, Z_i the i :th break point, $i = 1, 2, 3$. G denotes the lump-sum transfer, and $G^{np,p}$ denotes the (inter-group) transfer.

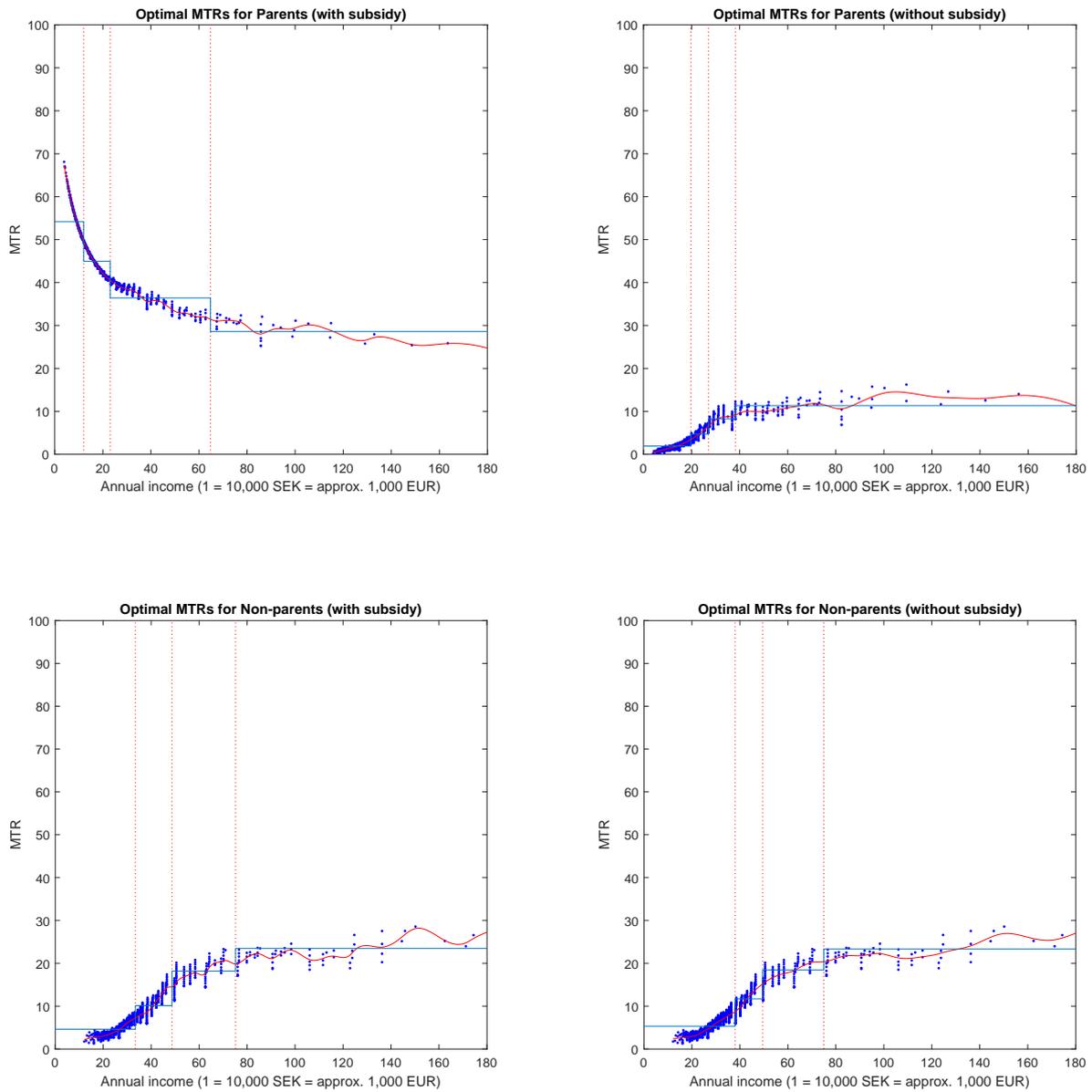


Figure 3: Optimal marginal tax rates for the fully nonlinear and four segment piecewise linear tax systems (Utilitarian with log transformation).

Table D.2: Welfare gain comparison (Utilitarian SWF with log transformation)

Optimum	Welfare Gain
Linear ($s = 0$)	Benchmark
Linear ($s = 0.76$)	$\approx 0\%$
<i>Difference</i>	$\approx 0\%$
Two bracket piece-wise linear ($s = 0$)	0.26%
Two bracket piece-wise linear ($s = 1$)	0.30%
<i>Difference</i>	0.04%
Four bracket piece-wise linear ($s = 0$)	0.29%
Four bracket piece-wise linear ($s = 1$)	0.34%
<i>Difference</i>	0.05%
Fully Nonlinear ($s = 0$)	0.44%
Fully Nonlinear ($s = 1$)	0.49%
<i>Difference</i>	0.05%