CESifo Working Papers

www.cesifo.org/wp

Public Provision of Private Goods, Tagging and Optimal Income Taxation with Heterogeneity in Needs

Spencer Bastani
Sören Blomquist
Luca Micheletto

CESIFO WORKING PAPER NO. 3275
CATEGORY 1: PUBLIC FINANCE
DECEMBER 2010

CESifo *Working Paper No. 3275*

Public Provision of Private Goods, Tagging and Optimal Income Taxation with Heterogeneity in Needs

Abstract

Previous literature has shown that public provision of private goods can be a welfare-enhancing device in second-best settings where governments pursue redistributive goals. However, three issues have so far been neglected. First, the case for supplementing an optimal nonlinear income tax with public provision of private goods has been made in models where agents differ only in terms of market ability. Second, the magnitude of the welfare gains achievable through public provision schemes has not been assessed. Third, the similarities/differences between public provision schemes and tagging schemes have not been thoroughly analyzed. Our purpose in this paper is therefore threefold: first, to extend previous contributions by incorporating in the theoretical analysis both heterogeneity in market ability and in the need for the publicly provided good; second, to perform numerical simulations to quantify the size of the potential welfare gains achievable by introducing a public provision scheme, and to characterize the conditions under which these welfare gains are sizeable; finally, to compare the welfare gains from public provision with the welfare gains from tagging.

JEL-Code: H21, H42.

Keywords: optimal income taxation, in-kind transfers, tagging.

Spencer Bastani
Uppsala Center for Fiscal Studies at the Department of Economics
Uppsala University
Sweden – 75120 Uppsala
spencer.bastani@nek.uu.se

Sören Blomquist
Uppsala Center for Fiscal Studies at the Department of Economics
Uppsala University
Sweden – 75120 Uppsala
soren.blomquist@nek.uu.se

Luca Micheletto
Department of Institutional Analysis and Public Management
L. Bocconi University
Italy – 20136 Milan
luca.micheletto@unibocconi.it

We would like to thank for their comments Wojciech Kopczuk as well as seminar participants at University of Umeå, University of Urbana-Champaign, the PET 2009 conference in Galway, the IIPF 2009 conference in Cape Town, the NTA 2009 conference in Denver, the SIE 2009 conference in Rome.

1. Introduction

For a long time the conventional wisdom among economists has maintained that only public goods should be publicly provided and that cash transfers dominate in-kind transfers as a vehicle to attain redistributive goals. The latter prescription was grounded on the observation that, while in-kind transfers constrain the behavior of their recipients, cash transfers do not. However, virtually all governments conduct a significant amount of redistribution through in-kind programs and provide a large variety of goods that are basically private in nature.[1] A variety of reasons has over time been put forward to rationalize this preference for redistributing in-kind rather than in cash.[2] Our paper relates to what can be labelled the "self-targeting" justification for in-kind transfers, namely the idea that, in a second-best setting, they may serve as a screening device between those who are the intended beneficiaries of the government's redistributive policy and those who are not. Such a rationale for in-kind transfers was already contained in the contributions by Nichols and Zeckhauser (1982) and Guesnerie and Roberts (1984), but it is only since the beginning of the nineties that scholars have started to incorporate public provision of private goods into an otherwise standard optimal nonlinear income taxation model.[3] These more recent contributions have shown that public provision (hereafter simply PP) of private goods can alleviate the informational problems that restrict redistribution when the identity of high- and low-skilled agents is hidden and the only information available to the government pertains to the distribution of abilities in the population and the form of the individuals' utility function.[4]

There are three issues that have not yet been addressed in the literature and that we intend to consider in this paper. The first is that the case for supplementing an optimal nonlinear income tax with PP of private goods has been made in models where agents differ only in terms of market ability. However, if one thinks of some of the examples of welfare-enhancing publicly provided goods that are typically cited in the literature, as for instance child-care services or elderly-care services, it is apparent that

[1]Public provision of private goods often represents an amount close to 20% of GDP.

[2]See Currie and Gahvari (2008) for a recent survey.

[3]Throughout the paper we use the expressions "in-kind transfers" and "public provision of private goods" as synonimous.

[4]See e.g. Blomquist and Christiansen (1995), Boadway and Marchand (1995), Cremer and Gahvari (1997), Balestrino (2000) Pirttil and Tuomala (2002) and Micheletto (2004).

not all agents need the publicly provided good. Since the results derived in previous contributions do not necessarily carry over in a simple way to models where agents differ along several dimensions, it is important to investigate the effects of in-kind transfers in models with heterogeneity both in skills and in needs for the publicly provided good. Another issue that has been neglected in the literature is the quantitative assessment of the welfare gains achievable through PP of private goods. Previous contributions have been confined to theoretical modelling; it is still an open question whether the welfare gains from PP of private goods are significant or negligible. Finally, the literature has so far failed to recognize that there are some interesting similarities/differences between PP schemes and tagging schemes.

In this paper we provide a comprehensive assessment of the welfare gains achievable by PP in a model where (realistically) only a fraction of individuals need the publicly provided good. Moreover, we investigate whether the presence of a general nonlinear income tax is crucial for public provision to be welfare enhancing, or if welfare gains of similar magnitudes are achievable with linear or piece-wise linear tax schedules like those observed in real world economies. We also compare PP with tagging: both instruments relieve informational frictions but act on different self-selection constraints; it is therefore interesting to quantitatively assess which instrument is the more powerful redistributive device in realistically calibrated economies. One reason why we find this comparison particularly interesting is that tagging- and PP schemes are likely to have very different political appeal. Indeed, one of the key messages of our paper is that, as an instrument for targeting needy subgroups in the population, PP can serve in many relevant cases as a valid and politically viable substitute for tagging schemes which are likely to be deemed politically unfeasible.

We will use two different models in our analysis. The first is an implementation of an extended discrete-type optimal taxation model of the kind analyzed in Blomquist et al. (2010), which in turn builds on the Stern (1982) and Stiglitz (1982) two-type version of the Mirrlees' (1971) optimal nonlinear taxation model. The second model employs an empirically relevant labor supply specification and employs a large sample of discrete taxpayer types derived from the Swedish population distribution. For this model we derive optimal linear and piece-wise linear tax schedules like in Stern (1976) and Slemrod et al. (1994) respectively. To our knowledge the present paper is the first to analyze PP or tagging together with piece-wise linear tax schedules.

The results from our calibrated model indicate that the welfare gains from PP can be substantial, at least if the policy maker has access to a general nonlinear tax. Sizeable welfare gains are also achievable under a (4-brackets)-piece-wise linear tax, but are very small under a linear tax. Moreover, tagging and PP are almost equivalent, in terms of welfare gains, under an unrestricted nonlinear income tax or a 4-brackets-piecewise-linear income tax.

The paper is organized as follows. In section 2 we first present a simple theoretical optimal nonlinear income taxation model where agents differ both in skills and in needs for the publicly provided good. We then proceed by discussing how the pure income tax optimum can be improved upon by using a PP scheme, and we contrast the effects of a PP scheme with those descending from tagging. Section 3 presents the results from the numerical simulations performed on the stylized model,including an estimate of the welfare gains achievable by alternatively using PP or tagging. Section 4 develops our empirically driven simulation approach and extends the analysis of the effects of PP and tagging to settings where the income tax at disposal of the government is less sophisticated (linear or piecewise-linear) than the one assumed in section 2. Finally, section 5 offers concluding remarks.

2. The model

Our theoretical model is an extension of the one considered in a recent contribution by Blomquist et al. (2010). In that paper the authors analyzed the desirability of a PP scheme that had not been previously addressed in the literature, and they did so in an optimal nonlinear income taxation model where agents were assumed to differ only in terms of skills. The novelty of the PP scheme was that the publicly provided good was delivered free of charge and public budget balance was attained by adjusting the shape of the income tax. Another feature of their model was that the good x which was candidate for PP was a good that didn't enter the agents' utility function directly, but had to be acquired in order to work. In particular, for all agents the demand for the good was related to the hours of work through the same monotonically increasing function f, so that $x = f(h)$, where h denoted hours of work

Extending the Blomquist et al. (2010) paper to a bi-dimensional setting, we focus for illustrative purposes on child care services as a candidate for public provision. Accordingly, we regard agents as differing in terms of both labor productivity (wage rates) and need for child care services. Those who need child care services in order to work are for simplicity labelled "parents" (rather than "parents with children in child care age"). It is important to emphasize right from the outset that what we have in mind as an empirical counterpart of this label 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 this sense they can be singled out as the "users" of the publicly provided private good on which we focus. Thus, albeit in the paper we refer for simplicity to "parents" and "non-parents", one has to bear in mind that for our purposes the group of parents represents only a subset of the parents with young children, and that this subset consists to a large extent of mothers of young children.

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. Each agent chooses how much labor to supply and the corresponding consumption level, which depends on the tax liability.

Throughout the paper, we use the expression "agents of ability (or skill) type i" to refer to agents earning a wage rate w^i. There are in the population n different ability types of agents ordered in such a way that $w^1 < w^2 < ... < w^n$. The population size is normalized to unity, the proportion of agents of ability type i is π^i, and $\delta^i \in [0, 1]$ denotes the proportion of parents among agents of ability type i. The government knows both π^i and δ^i for all $i \in \{1, ..., n\}$. 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.[5] Child care

[5]This assumption is made for simplicity and it does not affect the qualitative results.

services do not represent a good that enters the parents' utility function directly; it entails for them 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 $\overline{w} \equiv w - q$ and w for, respectively, parents and non-parents. All agents have identical preferences over consumption and hours of work, represented by the utility function $u(c, h)$.[6]

2.1. A pure income tax optimum

Let's start with a characterization of the solution to the government's problem in the absence of public provision. The government's objective is to maximize a weighted sum of agents' utilities. For this purpose, the government has at its disposal a nonlinear income tax $T(Y)$. Based on the link between pre-tax earnings and post-tax earnings implied by the tax schedule, agents choose labor supply to maximize 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 $B = Y - T(Y)$ 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 to that intended for some other agent. An agent that chooses a bundle intended for someone else is called "mimicker".

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 $V^{i,p}(B, Y) = u\left(B - qY/w^i, Y/w^i\right)$ and $V^{i,np}(B, Y) = u\left(B, Y/w^i\right)$, for respectively parents and non-parents of ability type i. As customary in asymmetric information optimal taxation models, the relative slope of the indifference curves of various agents in the (Y, B)-space plays an important role. In models where agents only differ along the wage rate dimension, (weak) normality of consumption is a sufficient condition to ensure that the agent monotonicity property holds. This property implies that, at any given point in the (Y, B)-space, the indifference curves are flatter the higher

[6] In this paper we disregard other possible costs or benefits of having children.

the wage rate of an agent. In our model where agents also differ according to whether they are parents or not, one might be tempted to conjecture that, in a setting without public provision of child care services, a similar property holds once the wage rates for parents are considered *net* of the unitary cost of child care services. This conjecture is however wrong, at least for a general utility function $u(c,h)$. To see this, consider a parent earning a unitary wage rate w^p and a non-parent earning a unitary wage rate $w^{np} = w^p - q$. If the conjecture were correct, these two agents should have equally sloped indifference curves at any bundle in the (Y,B)-space.[7] However, if we calculate the marginal rates of substitution between Y and B for the two agents at the same (Y,B)-bundle, we get

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

for the parent agent, whereas

$$MRS^{np} = \frac{1}{w^p - q}\left[-\frac{\partial u\left(B, \frac{Y}{w^p-q}\right)/\partial h}{\partial u\left(B, \frac{Y}{w^p-q}\right)/\partial c}\right] \tag{2}$$

for the non-parent agent. It is easy to recognize that there is no reason to expect that the values of (1) and (2) coincide.

The fact that the agent monotonicity property might not hold in our setting implies that the single-crossing property might be violated too. The solution to the government's problem might therefore not satisfy the monotonic chain to the left property.[8] For this reason we write the government's problem incorporating all the possible self-selection constraints.

Denoting by $\alpha^{i,p}$ the welfare weight used by the government for parents of ability type i $(i = 1, ..., n)$ and by $\alpha^{i,np}$ the welfare weight applied to non-parents of ability type

[7] Notice that in a *laissez faire* equilibrium a parent with wage rate w^p and a non-parent with wage rate $w^{np} = w^p - q$ would behave identically, and get the same level of utility, since their preferences in the (h,c)-space, as well as their budget set, are the same. In a *laissez faire* equilibrium a parent earning a unitary wage rate w^p would maximize $u(c,h)$ subject to the budget constraint $c = w^p h - qh$ whereas a non-parent earning a unitary wage rate w^{np} would maximize $u(c,h)$ subject to the budget constraint $c = w^{np}h$. When $w^{np} = w^p - q$ the two problems, and their solution, coincide.

[8] This property is satisfied when the only binding self-selection constraints are those running downwards and linking pair of adjacent types. For further details, see Guesnerie and Seade (1982).

i ($i = 1...,n$), with $\sum_{i=1}^{n}\alpha^{i,p} + \sum_{i=1}^{n}\alpha^{i,np} = 1$, the optimal taxation problem (hereafter OT problem) solved by the government can be formally written as:

$$\max_{\{B^{i,s},Y^{i,s}\}} \quad \sum_{s=p,np}\sum_{i=1}^{n}\alpha^{i,s}V^{i,s}\left(B^{i,s},Y^{i,s}\right)$$

subject to:

$$\begin{aligned}
V^{i,p}(B^{i,p},Y^{i,p}) &\geq V^{i,p}(B^{j,p},Y^{j,p}), \quad i,j \in \{1,...,n\}, i \neq j && (\lambda^{i,p;j,p}) \\
V^{i,p}\left(B^{i,p},Y^{i,p}\right) &\geq V^{i,p}(B^{j,np},Y^{j,np}), \quad i,j \in \{1,...,n\}, && (\lambda^{i,p;j,np}) \\
V^{i,np}(B^{i,np},Y^{i,np}) &\geq V^{i,np}(B^{j,np},Y^{j,np}), \quad i,j \in \{1,...,n\}, i \neq j && (\lambda^{i,np;j,np}) \\
V^{i,np}(B^{i,np},Y^{i,np}) &\geq V^{i,np}(B^{j,p},Y^{j,p}), \quad i,j \in \{1,...,n\}, && (\lambda^{i,np;j,p})
\end{aligned}$$

$$\sum_{i=1}^{n}\pi^{i}\left[(Y^{i,p}-B^{i,p})\delta^{i}+(Y^{i,np}-B^{i,np})(1-\delta^{i})\right] \geq 0, (\mu)$$

where Lagrange multipliers are within parentheses, the first four sets of constraints represent the self-selection constraints, and the last constraint is the government's budget constraint.

The first set of self-selection constraints ($\lambda^{i,p;j,p}$-constraints) incorporates all the possible constraints linking a parent of a given ability type to a parent of a different ability type. The second set ($\lambda^{i,p;j,np}$-constraints) contains all the possible constraints requiring that a parent of a given ability type should not be tempted to choose a bundle intended for a non-parent. For simplicity, we can label these two sets of constraints as respectively the parent/parent self-selection constraints and the parent/non-parent self-selection constraints. Adapting this terminology to the last two sets of incentive constraints, we can see that the third set ($\lambda^{i,np;j,np}$-constraints) contains all the possible non-parent/non-parent self-selection constraints and the fourth set ($\lambda^{i,np;j,p}$-constraints) all the possible non-parent/parent self-selection constraints.

Grouping the self-selection constraints in this way is convenient for the purpose of highlighting the welfare properties of PP. As we will soon explain in more details, the reason is that PP alleviates only some of the self-selection constraints, whereas it may have no effect or even exacerbate some other self-selection constraints.

Presenting the self-selection constraints as above is also useful to illustrate the relative merits of PP- versus tagging schemes. Anticipating what will be clarified below, the reason is that these two schemes interact in a different way with the different sets of self-selection constraints.

Intuition therefore suggests that whether a PP scheme outperforms a tagging scheme or not depends on which self-selection constraints represent for the government, at the solution to the OT problem, the most severe obstacle in achieving the redistributive goals.

Manipulating the first order conditions of the OT problem above, one can easily show that the general expression for the marginal tax rate faced by a parent of ability type i is given by:

$$\begin{aligned} T'\left(Y^{i,p}\right) &= \frac{1}{\mu\pi^{i}\delta^{i}}\left[\sum_{j\neq i}\lambda^{j,p;i,p}V_{B}^{j,p;i,p}\left(MRS^{i,p}-MRS^{j,p;i,p}\right)\right]+ \\ & \frac{1}{\mu\pi^{i}\delta^{i}}\left[\sum_{j=1}^{n}\lambda^{j,np;i,p}V_{B}^{j,np;i,p}\left(MRS^{i,p}-MRS^{j,np;i,p}\right)\right], \end{aligned} \quad (3)$$

whereas the general expression for the marginal tax rate faced by a non-parent of ability type i is given by:

$$\begin{aligned} T'\left(Y^{i,np}\right) &= \frac{1}{\mu\pi^{i}\left(1-\delta^{i}\right)}\left[\sum_{j\neq i}\lambda^{j,np;i,np}V_{B}^{j,np;i,np}\left(MRS^{i,np}-MRS^{j,np;i,np}\right)\right]+ \\ & \frac{1}{\mu\pi^{i}\left(1-\delta^{i}\right)}\left[\sum_{j=1}^{n}\lambda^{j,p;i,np}V_{B}^{j,p;i,np}\left(MRS^{i,np}-MRS^{j,p;i,np}\right)\right]. \end{aligned} \quad (4)$$

The results provided by (3)-(4) are quite standard and we do not discuss them at length. The only reason to distort agents' (labor supply) behavior is the presence of binding self-selection constraints. Eq. (3) tells us that if there are agents, other than parents of ability type i, who are indifferent between choosing the bundle intended for them and the bundle intended for parents of ability type i, the labor supply of the latter has to be distorted to deter mimicking. Similarly, eq. (4) tells us that if there are agents, other than non-parents of ability type i, who are indifferent between choosing the bundle intended for them and the bundle intended for non-parents of ability type i, the labor supply of the latter has to be distorted to prevent mimicking.

Notice that, in contrast to what happens in standard OT models, we cannot rule out the possibility that for some of the marginal income tax rates, as defined by (3)-(4), more than one self-selection constraint is binding at the same time. In particular, the monotonic-chain-to-the-left property is no longer necessarily satisfied since, as we previously pointed out, the agent monotonicity assumption is likely to be violated.

Let's consider now what happens when public provision of child care services is introduced.

2.2. An optimum with public provision

Following Blomquist et al. (2010) we consider a PP scheme which is entirely financed through the income tax and where people can get free of charge as much child care services as they need.[9] In such a setting all agents, irrespective of whether they are parents or not, have at any given point in the (Y, B)-space an indirect utility which is given by $V^i\left(B^i, Y^i\right) = u\left(B^i, Y^i/w^i\right)$.[10] In other words, child care purchases no longer appear in the (private) budget constraints of parents but instead enter the government's budget constraint. Defining β^i, for $i = 1, ..., n$ as $\beta^i \equiv \alpha^{i,p} + \alpha^{i,np}$, the income-tax-cum-public-provision problem (hereafter PP problem) solved by the government can be formally written as:

$$\max_{\{B^i, Y^i\}} \quad \sum_{i=1}^{n} \beta^i V^i\left(B^i, Y^i\right)$$

[9] Notice that our assumptions imply that the demand for child care is characterized by satiation conditional on labor supply. Without satiation, it would not be possible to offer any amount free of charge as agents would expand their consumption beyond any reasonable limit, unless some private disutility (time costs, etc.) is incurred in order to consume the publicly provided good.

[10] Therefore, as compared to the case considered in the previous section, 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 PP scheme provided that the income effects are not very large.

subject to:

$$V^i\left(B^i, Y^i\right) \geq V^i\left(B^j, Y^j\right), \qquad j \neq i, \quad i, j = 1, ..., n, \qquad \left(\lambda^{ij}\right)$$

$$\sum_{i=1}^{n} \pi^i\left(Y^i - B^i\right) \geq q \sum_{i=1}^{n} \pi^i \delta^i \frac{Y^i}{w^i}, \qquad (\mu)$$

where the presence of a single set of self-selection constraints reflects the fact that with PP the agents' behavior only depends on their ability type and no longer on their parental status. Before characterizing analytically the expressions for the marginal tax rates faced by the various agents under an income-tax-cum-public-provision optimum, let's first consider how the introduction of PP is going to affect the self-selection constraints faced by the government. Previous contributions have pointed out how, when agents differ only in terms of ability and the government aims at redistributing from the high-skilled to the low-skilled, the PP of a complementary-to-labor good allows achieving a Pareto-improvement upon the pure income tax optimum.[11] The logic of the argument can be easily illustrated by referring to our model above. The analysis is simplified by analyzing separately the different patterns of self-selection constraints which can arise in the economy.

2.2.1. Disjoint wage supports

Suppose all agents of ability types $k \in \{1, ..., i\}$ are parents and all agents of ability types $z \in \{i+1, ..., n\}$ are non-parents. Suppose further that a fully separating equilibrium with $Y^1 < ... < Y^n$ is achieved as the solution to the government's OT problem. Finally, suppose that at the solution to the OT problem the only binding self-selection constraints run downwards from higher ability agents to lower ability agents. To show that a Pareto-improvement can be obtained by using PP one can proceed as follows. Denote by $\left(Y^{j*}, B^{j*}\right)$ the bundle offered to agents of skill type $j = 1, ..., n$ at the solution to the OT problem. Let agents get the amount of child care services that they want and, instead of the original set of bundles $\left(Y^{j*}, B^{j*}\right)$, offer the following packages:

[11] See Blomquist et al. (2010).

$\left(Y^{1*}, B^{1*} - qY^{1*}/w^1\right)$,..., $\left(Y^{i*}, B^{i*} - qY^{i*}/w^i\right)$, $\left(Y^{i+1*}, B^{i+1*}\right)$,..., (Y^{n*}, B^{n*}). 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 tax payment of each type of parents has been increased just enough to cover the cost of publicly provided services demanded by that type (qY^{j*}/w^j for $j = 1, ...i$). The only effects of the reform that are left to evaluate are those on the binding self-selection constraints.

Consider first the self-selection constraints requiring higher ability non-parents to be prevented from mimicking lower ability non-parents. Clearly, these self-selection constraints are unaffected by the proposed reform. Consider now the self-selection constraint requiring a non-parent of ability type $i + 1$ to be prevented from mimicking a parent of ability type i. By assumption, this was the only binding self-selection constraint linking parents and non-parents at the solution to the OT problem. It should be apparent that this self-selection constraint is weakened by the proposed reform. The reason is that the consumption that a type $i + 1$ agent can get by mimicking a type i agent is now lower (by the amount qY^{i*}/w^i) than before the reform, whereas the labor effort that he/she has to exert has not changed. A similar reasoning leads to the conclusion that also all the self-selection constraints requiring higher ability parents to be prevented from mimicking lower ability parents are relaxed. The consumption that a parent of skill type $j - 1$ can get by mimicking a parent of skill type j is now lower (by the amount $q\left[Y^{j-1*}/w^{j-1} - Y^{j-1*}/w^j\right]$) than before the reform, whereas the labor effort that he/she has to exert has not changed. We can therefore conclude that PP is an unambiguously welfare-enhancing instrument in this case. We can also notice, by comparing the extra burden that it imposes on a type $i+1$ mimicker as compared to the extra burden imposed on a type $j \in \{2, ..., i\}$ mimicker, that it is an especially effective policy instrument when it comes to deterring non-parents from mimicking parents.[12]

[12]The extra burden placed on skill type $i + 1$ mimickers (who are non-parents) depends only on the labor supply of the mimicked agents (who are the parents of skill type i). The extra burden placed on mimickers of skill type $j \in \{2, ...i\}$ (who are parents) depends instead on the positive difference between the labor supply of the mimicked and the labor supply of the mimicker. The smaller this difference, the smaller the extra burden placed on the mimicker by the introduction of PP.

2.2.2. Overlapping wage supports

Let's now modify slightly our example and introduce overlapping wage supports for parents and non-parents. For this purpose, let's assume that n, the number of different ability types in the economy, is an even number and that for $i \in \{1, 2, 3, ..., n/2\}$ all agents of ability type $k = 2i - 1$ are parents whereas all agents of ability type $z = 2i$ are non-parents. Suppose again that a fully separating equilibrium with $Y^1 < ... < Y^n$ is achieved as the solution to the government's OT problem and that the only binding self-selection constraints run downwards and link pair of adjacent types. What would be the effects of introducing PP in a set-up like this? To answer this question consider the effects of a reform shaped along the same lines as above, and denote by $\left(Y^{j**}, B^{j**}\right)$ the bundle offered to agents of type $j = 1, ..., n$ at the solution to the OT problem in the absence of PP. As before, let agents get the amount of child care services that they want and, instead of the original set of bundles $\left(Y^{j**}, B^{j**}\right)$, offer the following packages: $\left(Y^{k**}, B^{k**} - qY^{k**}/w^k\right)$ and (Y^{z**}, B^{z**}) where $i \in \{1, 2, 3, ..., n/2\}$, $k = 2i - 1$ and $z = 2i$. Once again, 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. Consider any of the self–selection constraints requiring a non-parent of ability type z to be prevented from mimicking a parent of ability type $z - 1$. These constraints are weakened since the consumption that a type z agent can get by mimicking a type $z - 1$ agent is now lower (by the amount qY^{z-1**}/w^{z-1}) than before the reform, whereas the labor effort that he/she has to exert has not changed.

So far the analysis has confirmed the virtue of PP as an instrument to alleviate the incentive problems faced by the government. Moreover, since by assumption we have $n/2$ self-selection constraints involving non-parents being tempted to mimic parents, the effect of the introduction of PP is potentially quite strong. However, we need to evaluate the effects produced on the remaining $(n/2) - 1$ binding self-selection constraints, those requiring parents of ability type k to be prevented from mimicking non-parents of ability type $k - 1$.[13] In this case the proposed reform has a perverse effect on the mimicking incentives. In fact, at the pre-reform equilibrium a parent of ability type k was assumed

[13] The number of binding self-selection constraints is in this case equal to $(n/2) - 1$ due to the fact that, according to our assumptions, parents of ability type 1 are not tempted to mimic anyone else.

to be indifferent between choosing the bundle intended for him/her (Y^{k**}, B^{k**}) and the one intended for a non-parent of ability type $k-1$ (Y^{k-1**}, B^{k-1**}). After the reform, his/her utility as a non-mimicker would be unchanged: he/she would still work for Y^{k**}/w^k hours and consume $B^{k**} - qY^{k**}/w^k$. His/her utility as a mimicker, on the other hand, would increase: mimicking a type $k-1$ non-parent would still require to work for Y^{k-1**}/w^k hours, as before the reform, but consumption would increase from $B^{k-1**} - qY^{k-1**}/w^k$ to B^{k-1**}. Thus, the proposed reform has a detrimental effect on all the binding self-selection constraints requiring parents of ability type k to be prevented from mimicking non-parents of ability type $k-1$. This means that it is no longer obvious that PP can be used to accomplish a Pareto improving reform.

2.2.3. Intersecting wage supports

Finally, let's abandon the assumption that for each given ability type either all agents are parents or all agents are non-parents. Then, another effect of the PP scheme is that at the solution to the income-tax-cum-public-provision problem the government is forced to pool all agents of a given ability type, irrespective of their parental status. The reason is apparent. With the introduction of the PP scheme the shape of the agents' indifference curves in the (Y, B)-space only depends on the agents' gross wage rate w; parental status becomes irrelevant. Thus, parents and non-parents, provided that they are of the same ability type, become indiscernible with PP in the sense that they cannot be separated by properly designing the nonlinear income tax. It is however worth pointing out that PP might in some other cases help the government to screen between groups of agents that would be pooled under a pure income tax optimum. For instance, if a parent has a net hourly wage rate $w-q$ which is close enough to the hourly wage rate of a non-parent, it could be the case that the government pool them together at the solution to the OT problem. With the PP scheme in place, on the other hand, it would be easier for the government to separate the two groups of agents.

2.2.4. Optimal marginal tax rates

We are now ready to characterize optimal marginal tax rates. Manipulating the first order conditions of the government's problem, the general expression for the marginal tax rate faced by an agent of ability type i is given by:

$$T'\left(Y^i\right) = \frac{1}{\mu\pi^i}\left[\sum_{j\neq i}\lambda^{ji}V_B^{ji}\left(MRS^i - MRS^{ji}\right)\right] + \delta^i\frac{q}{w^i}. \tag{5}$$

Given that the agent monotonicity assumption is recovered with the introduction of PP, and under the reasonable assumption that the government aims at redistributing from the higher ability types to the lower ability types, the solution to the PP problem entails a simple monotonic chain to the left. We would therefore have

$$T'\left(Y^n\right) = \delta^n\frac{q}{w^n}, \tag{6}$$

whereas for $i = 1, ..., n-1$ and $j = i+1$ we would have

$$T'\left(Y^i\right) = \frac{\lambda^{ji}V_B^{ji}}{\mu\pi^i}\left(MRS^i - MRS^{ji}\right) + \delta^i\frac{q}{w^i}. \tag{7}$$

A special case of (6)-(7) occurs when, for any given ability type, either all agents are parents or all agents are non-parents. The government would then not be forced to pool some parents with some non-parents and offer them the same allocation in the (Y, B)-space. An equilibrium which is fully separable in this sense is in principle attainable and the deltas appearing in the equations (6)-(7) would either take a value equal to 0 (for the ability types represented solely by non-parents) or equal to 1 (for the ability types represented solely by parents). If this is the case, the expressions defining the marginal tax rates faced by the various non-parents only incorporate a self-selection term, whereas those defining the marginal tax rates for the various parents incorporate a self-selection term plus a q/w term, with w depending on the specific wage rate of the parents under consideration.

An interesting thing to notice about (6)-(7) when $\delta^i \in \{0, 1\}$ $\forall i \in \{1, ..., n\}$ is that, even if the introduction of PP is likely to lead to an increase in some of the marginal tax rates (those for parents, due to the presence of the q/w terms), total distortions in the economy may 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, as emphasized in Blomquist et al. (2010), serve the same role as a market price in letting parents face the right incentives.[14] On the other hand, PP serves the purpose of weakening some of the binding self-selection constraints. For these constraints the mimicking-deterring-effect makes less urgent the need to distort agents for self-selection purposes and therefore allows reducing the truly distortionary component (i.e. the λ-terms) in the formulas for the marginal tax rates. Notice also that in this fully separating equilibrium the expressions for the marginal tax rates that apply to non-parents do not incorporate the q/w terms. This is important since these terms would for them represent a truly distortionary component.[15]

In the general case when δ^i is not restricted to be either equal to 0 or 1, PP forces the government to offer the same (Y, B)-bundle to both parents and non-parents of a same ability type i. With respect to the marginal tax rates, this partial pooling feature of the solution to the PP problem has the unappealing consequence that the distortion faced by the non-parents of ability type i is likely to be exacerbated by the introduction of PP. The reason is that their marginal tax rate also incorporates a term that reflects the marginal resource cost of the publicly provided service used by the parents of ability type i with whom they are pooled. Raising the marginal tax rate by q/w^i is non-distortive and fully corrective for type i parents but it is a fully distortive term for type i non-parents. The government is therefore faced with the trade-off between raising the distortion on type i non-parents and correcting the distortion imposed on type i parents by the introduction of the free of charge PP scheme. According to (6)-(7) this trade-off is solved raising the value of the marginal tax rate by an amount which is smaller than the one required to fully correct the behavior of type i parents.[16] In particular, as the proportion of parents among type i agents becomes larger and larger, the increase in the marginal tax rate becomes closer and closer to the one required to fully correct the behavior of type i parents. Finally, notice that for non-parents the marginal tax rates defined by (6)-(7) are entirely distortionary, whereas for parents the

[14] It 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.

[15] Notice however 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 PP of child care are raised only from parents. It means that if non-parents participate too in the financing of the PP of child care services, the additional revenue extracted from them may to a large extent be collected in a non-distortionary way through an increase in inframarginal tax rates.

[16] From this point of view we can say that type i parents are induced to work too much and type i non-parents too little.

truly distortionary component is given by $-(1-\delta^n)\, q/w^n$ in the case of (6) and by $\left(\lambda^{jz} V_B^{jz}/\mu\pi^z\right)\left(MRS^z - MRS^{jz}\right) - (1-\delta^z)\, q/w^z$ in the case of (7).

After having considered the effects of PP, in the next section we briefly discuss the case where the government supplements income taxation with a tagging scheme.

2.3. An optimum with tagging

The basic distinguishing mechanism in the tagging problem as compared to the OT problem outlined in section 2.1 is that with tagging the population may be disaggregated into groups, using some observable characteristic or "tag", to each of whom a different nonlinear income tax schedule applies.[17] The government then solves a set of separate optimal income tax problems, one for each of the tagged groups, with the possibility of accomplishing lump-sum inter-group transfers. Taking as a benchmark the solution to the OT problem, the welfare-enhancing potential of a tagging scheme comes from the possibility to eliminate all self-selection constraints linking agents belonging to two separate tagged groups.

In our model with parents and non-parents, one could view parental status as a possible tag that allows disaggregating the population into two separate groups with two distinct tax schedules applying to parents and non-parents. From a formal point of view, however, we can still set up the government's problem in the way we did for the OT problem presented in section 2.1, the only difference being that we can now neglect all the parent/non-parent- and the non-parent/parent self-selection constraints since they will necessarily be slack. The income-tax-cum-tagging problem (hereafter TG problem) can therefore be written taking the OT problem of section 2.1 and setting $\lambda^{i,np;j,p} = \lambda^{i,p;j,np} = 0$.

Manipulating the first order conditions of the TG problem, the general expression for the marginal tax rate faced by a parent of ability type i is given by:

$$T'\left(Y^{i,p}\right) = \frac{1}{\mu\pi^i\delta^i}\left[\sum_{j\neq i}\lambda^{j,p;i,p} V_B^{j,p;i,p}\left(MRS^{i,p} - MRS^{j,p;i,p}\right)\right],$$

[17] The 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) Cremer et al. (2010).

whereas that for the marginal tax rate faced by a non-parent of ability type i is given by:

$$T'\left(Y^{i,np}\right)=\frac{1}{\mu\pi^{i}\left(1-\delta^{i}\right)}\left[\sum_{j\neq i}\lambda^{j,np;i,np}V_{B}^{j,np;i,np}\left(MRS^{i,np}-MRS^{j,np;i,np}\right)\right].$$

Before concluding this section and resorting to numerical simulations to assess the welfare properties of PP versus tagging, notice that the desirability of both PP and tagging hinges on the effects exerted on the self-selection constraints thwarting the government in pursuing its redistributive goals. There are however two main differences. The first is that PP acts on the incentive constraints both within- and across groups, whereas tagging prevents mimicking across groups but is ineffective in alleviating incentives problems within groups. The second difference is that PP might relax some incentive constraints and tighten some others, whereas tagging never tightens incentive constraints. This also implies that tagging, contrary to PP, always allows achieving a Pareto improvement upon the pure income tax optimum.[18]

3. The importance of the wage distribution overlap: numerical simulations

To provide additional insights, we perform numerical simulations on three variants of a simple model with just four different groups of agents. The purpose is to highlight that the effectiveness of PP and tagging in slackening self-selection constraints is affected both by the way users and non-users are distributed across ability types and by the redistributive tastes of the government. To assess the sensitivity of results on the degree of social aversion to inequality, we consider the two polar cases of a utilitarian social welfare function and of a max-min social welfare function. Regarding the distribution of users and non-users across ability types, we consider the following three cases:

[18] It is however also true that the introduction of a PP scheme always allows achieving a Pareto improvement upon the solution to the government's TG problem. The reason is that tagging eliminates all the self-selection constraints linking parents to non-parents and vice versa. PP would then only affect the self-selection constraints linking parents to parents and we know that in this case its effect would be beneficial, provided that the government aims at redistributing from the relatively well-off to the relatively worse-off.

Case 1 (disjoint). agents can be of four different ability types characterized by the wage rates $w^1 < w^2 < w^3 < w^4$; all agents of ability type 1 and 2 are parents whereas all agents of ability type 3 and 4 are non-parents;

Case 2 (overlapping). agents can be of four different ability types characterized by the wage rates $w^1 < w^2 < w^3 < w^4$; all agents of ability type 1 and 3 are parents whereas all agents of ability type 2 and 4 are non-parents;

Case 3 (pooling). agents can be of three different ability types characterized by the wage rates $w^1 < w^2 < w^3$; all agents of ability type 1 are parents and all agents of ability type 3 are non-parents; as regards agents of ability type 2, there are among them both parents and non-parents.

3.1. Parameterization

In this section we employ the following utility function commonly used in the optimal taxation literature

$$u(c, h) = \log c - \frac{h^{1+k}}{1+k},$$

where we set $k = 2$ implying a compensated elasticity of labor supply equal to 1/3. For both case 1 and case 2 the values of wages are $w^1 = 2$, $w^2 = 2.8$, $w^3 = 4.2$ and $w^4 = 6$, and the benchmark level of the price of child care is $q = 1$. For case 3 the values of wages are $w^1 = 1$, $w^2 = 2$ and $w^3 = 3$, and the benchmark level of q is 0.5.[19] For all cases we present results for economies where the proportion of parents is equal to 15% and 30%.[20] Since it is of interest to compare the effects of the introduction of a PP scheme with those produced by supplementing optimal income taxation with tagging, we also calculate the solution to the government's problem under the assumption that two separate nonlinear income tax schedules apply to parents and non-parents. Finally, we also investigate the combined effect of both PP and tagging.

[19]The values of q have been chosen so that the ratio q/w^1 approximately mimics empirically relevant magnitudes of the fraction of the wage spent on child-care services for low earners which has been shown to lie around 45% (See Blomquist et al. (2010)).

[20]Parents are always assumed to be equally distributed between low-skilled parents and high-skilled parents. The same is true for non-parents.

3.2. Results

The results of our simulations are presented in appendix A. Tables A1-A3 provide the results for the utilitarian social welfare function; Tables A4-A6 those for the max-min social welfare function. The "disjoint" case is considered in Tables A1 and A4, the "overlapping" case in Tables A2 and A5, and the "pooling" case in Tables A3 and A6. Each table presents the pre-tax labor income and marginal tax rate for the various agents under the pure income-tax optimum (OT), the income-tax-cum-public-provision optimum (PP), the tagging optimum (TG), and the optimum where tagging and public provision are jointly used (TG+PP). The welfare gains associated with each of the three deviations from the pure income tax optimum are reported at the bottom of the various tables.

With one exception, the introduction of PP always raises the marginal tax rates faced by parents.[21] However, taking into account that under the PP scheme part of the marginal tax rate faced by parents represents a non-distortionary component, the truly distortionary component is lower for parents under an optimum with PP than under a pure income tax optimum.[22] With respect to the low-skilled non-parents, we find that in case 1 (and to a lesser extent also in case 3) PP implies a reduction in the marginal tax rate, whereas in case 2 it entails an increase in the marginal tax rate, especially under a max-min social welfare function.[23] Relying on the discussion contained in section 2.2, the increase observed in case 2 can be interpreted as a consequence of the fact that the provision scheme tightens the self-selection constraint requiring type 3 agents (parents) not to mimic type 2 agents (non-parents).

In accordance with the effect on the equilibrium marginal tax rates, we find that the introduction of the PP scheme tends to boost the labor supply of parents. In general also

[21] The exception is represented by the marginal tax rates faced by the high-skilled parents in case 3 when the government maximizes a max-min objective function. Under the OT solution the high-skilled parents do not work and face a quite high marginal tax rate (75%). The marginal tax rate is lowered by the introduction of PP since it becomes optimal for the government to let high-skilled parents participate in the labor market.

[22] The only exception is represented by the high-skilled parents under a max-min social welfare function and an overlapping wage distribution. As we have discussed in section 2.2 the distortionary component of the marginal tax rate faced under a PP scheme by a parent earning a wage rate w is obtained by subtracting q/w from the "nominal" value of the marginal tax rate.

[23] For high-skilled non-parents the marginal tax rate is always zero, irrespective of whether there is PP or not.

the labor supply of non-parents tends to increase after the introduction of the PP scheme. The only relevant exception is in case 2 with the labor supply of low-skilled non-parents who, as we have already seen, are likely to face a significantly higher marginal tax rate under the PP regime when the social welfare function is of the max-min type.

As compared to the PP optimum, the main difference of the TG solution is that the marginal tax rate faced by high-skilled parents is always equal to zero since tagging allows achieving a within-groups no-distortion-at-the-top result. For the within-groups low-skilled agents, the TG optimum entails in cases 1 and 3 a reduction in the marginal tax rates as compared to the OT optimum. In case 2, however, low-skilled non-parents face, under the TG optimum, marginal tax rates that are higher than under the OT optimum.

To calculate a consumption-based measure of the welfare gains attainable by either the introduction of the PP scheme or by tagging, we consider an equivalent-variation-type of welfare gain measure. To obtain it, we proceed as follows. We look for the minimum amount of extra revenue that should be injected into the government's budget in the OT problem in order to achieve the same social welfare level under an OT optimum as under the PP optimum, the TG optimum or the TG+PP optimum.[24] Once we have found this minimum amount of extra revenue, we divide it by the aggregate consumption at the pure tax optimum in order to get a revenue-based measure of the welfare gains from PP, from tagging, or from the combined use of PP and tagging. We denote this revenue-based welfare gain measure by WG^r.

Another welfare gain measure that is often used in the literature is the one that is obtained by calculating the factor θ which, when multiplied with the consumption of all agents in the pre-reform equilibrium (which for us would be the OT equilibrium), achieves the same level of social welfare as under the post-reform equilibrium (which for us would be the PP, TG, or TG+PP equilibrium). The welfare gain measure is then

[24] Formally, let $E(T, SW)$ denote the smallest amount of "money from heaven" needed to achieve the social welfare level SW when the tax system is T. As our measure of the benfits descending from PP or tagging we respectively use $E(T^{OT}, SW(T^{PP})) - E(T^{OT}, SW(T^{OT})) = E(T^{OT}, SW(T^{PP}))$, $E(T^{OT}, SW(T^{TG})) - E(T^{OT}, SW(T^{OT})) = E(T^{OT}, SW(T^{TG}))$, and $E(T^{OT}, SW(T^{TG+PP})) - E(T^{OT}, SW(T^{OT})) = E(T^{OT}, SW(T^{TG+PP}))$, where T^{OT}, T^{PP}, T^{TG} and T^{TG+PP} denote respectively the optimal nonlinear tax in the absence of either PP or tagging, the optimal nonlinear-income-tax-cum-public-provision, the optimal nonlinear-income-tax-cum-tagging, and the optimal nonlinear income tax supplemented by both tagging and PP.

obtained as $\theta - 1$. If we had used this alternative way to measure welfare gains, the estimated benefits from either PP or tagging would have been larger than those that are reported in the tables below. The reason why we have not followed this approach is that if, starting at the OT equilibrium, one multiplies the consumption of all agents by a common factor, nothing guarantees that the self-selection constraints are still satisfied at the new allocation.[25]

Tables A5-A10 show that the introduction of PP always increases social welfare and that the welfare gains are increasing in the proportion of parents in the population. The welfare gains, measured as percentage of aggregate consumption, vary from a minimum of 0.68% to a maximum of 8.64%.

The results confirm that the welfare gains are sensitive to the type of self-selection constraints that are binding at a pure income tax optimum. Moreover, they seem to suggest that the beneficial effects of PP in alleviating the non-parent/parent self-selection constraints outweigh the detrimental effects of PP in reinforcing the parent/non-parent self-selection constraints. In fact, PP alleviates two self-selection constraints both in the "disjoint" case 1 and in the "overlapping" case 2, but in the latter it also tightens one self-selection constraint; nonetheless, PP delivers larger welfare gains in case 2 than in case 1. A rationale for this result can be found by noticing that the two self-selection constraints that are weakened by PP in case 2 are both of the non-parent/parent type, and we know that PP is very effective in alleviating this type of incentive constraints. In case 1, on the other hand, only one of the two self-selection constraints weakened by PP is of the non-parent/parent type; the other is of the parent-parent type, and we know that on these constraints the effectiveness of PP as a slackening device is lower.

The welfare gains displayed in Tables A1-A6 allows illustrating the effects of the interaction between the degree of social inequality to aversion and the degree of overlap between the wage distributions for parents and non-parents. In fact, an increase in the

[25]Notice that this happens despite the fact that the utility function is separable between leisure and consumption and logarithmic in consumption. The reason is the presence of both parents and non-parents. It implies that, whereas to multiply a non-parent's consumption by θ we need to multiply his/her after-tax income B by θ, to multiply a parent's consumption by θ we need to multiply his/her after-tax income B by a smaller number (equal to $q\left(Y/w\right)\left(\theta-1\right)/B$, where Y/w is the labor supply of the parent at the OT equilibrium). The fact that not all after-tax incomes are multiplied by the same factor is what determines the possibility that some self-selection constraints are violated. In particular, one can show that this is the case whenever the OT equilibrium is characterized by binding self-selection constraints where parents appear as potential mimickers.

social inequality to aversion has a sizeable effect on the welfare gains from PP only in cases 2 and 3. The reason has to do with the fact that a max-min social welfare function strengthens the importance of the self-selection constraint(s) preventing other agents from mimicking the least well-off members of the society.

In the “overlapping” case 2 this constraint is the one requiring low-skilled non-parents to be prevented from mimicking the low-skilled parents (who are the least well-off agents). Accordingly, we find that switching from a utilitarian to a max-min objective function more than doubles the welfare gains descending from the introduction of a PP scheme.

In the “disjoint” case 1 it is the high-skilled parents who have to be deterred from mimicking the least well-off agents (always the low-skilled parents). With a parent/parent self-selection constraint binding at the bottom of the income distribution, switching from a utilitarian to a max-min objective function has a minor effect on the welfare gains from PP.

The “pooling” case 3, where we have three ability types of agents and the intermediate type is represented by both parents and non-parents, is slightly more complex but the logic is still the same. Under a utilitarian objective function all agents of ability type 2 (parents and non-parents) are pooled together at the pure income tax optimum. Among the pooled agents, those who are parents have the steeper indifference curves. Thus, the tax function implementing the optimal allocation has to be designed to prevent two types of mimicking behaviors. On one hand, it has to prevent agents of ability type 3 from mimicking non-parents of ability type 2. On the other hand, it has to prevent parents of ability type 2 from mimicking agents of ability type 1. This represents a pattern of self-selection constraints very close to the one characterizing the pure income tax optimum in the “disjoint” case 1. Accordingly, the welfare gains from introducing PP are, under a utilitarian social welfare function, of the same magnitude as those obtained in the “disjoint” case 1. However, when the objective function is max-min, we can see from Table A6 that the solution to the OT problem entails bunching all parents at zero hours of work. In this case the tax function implementing the optimal allocation has to be designed to prevent non-parents of ability type 2 from mimicking parents of ability type 2 (besides preventing agents of ability type 3 from mimicking non-parents of ability type 2). This pattern of self-selection constraints is similar to the one characterizing the pure income tax optimum in the "overlapping" case 2. Accordingly, we find that the welfare gains from introducing PP in the "pooling" case 3 are quite large under a max-min objective function.

In the case of tagging, the possibility to apply different tax schedules to non-parents and parents implies a large transfer of resources from the former group to the latter with no need to worry about non-parents being tempted to mimic parents. Keeping this in mind and looking at the results displayed in Tables A1-A6, we can see that also tagging always entails an overall gain in terms of social welfare and that the welfare gains are increasing in the proportion of parents in the population. Expressed as a percentage of aggregate consumption, they vary from a minimum of 1.90% to a maximum of 8.54%. Even if the welfare gains from tagging tend to be larger than those from PP, it is interesting to note that there are cases when PP delivers larger welfare gains. This is for instance the case when the social welfare function is max-min and the distribution of parents and non-parents across skills is of the pooling type (case 3).

The results from the combined use of tagging and PP show that, once a tagging scheme is in place, the welfare gains achievable by introducing PP are quite small, ranging from 0% (case 1 under a utilitarian social welfare function) to about 0.9% (cases 2 and 3 under a max-min social welfare function).

Finally, even if not reported in the tables, we have also performed numerical simulations trying different values for the parameters q and k. Our results indicate that the welfare gains associated with PP tend to increase when q is increased and tend to decrease when k is increased (implying a lower compensated elasticity of labor supply).

4. An application to the Swedish economy

In this part of the paper we compare PP and tagging using an empirically driven simulation approach and considering various possibilities regarding the complexity of the income tax at disposal of the government.

In accordance with our remark at the beginning of section 2, in our simulation exercise we define parents as women with at least one child in day care age. Notice that this has a crucial implication for the interpretation of our tagging scheme. In particular, the tagging scheme is not to be interpreted as a scheme that assigns different tax schedules to families with young children and families without young children, but rather as a scheme that assigns a different tax schedule to women with at least one child in day care age.

To obtain measures of welfare gains from policy reforms achievable in real economies we make two modeling assumptions: (i) we choose an empirically relevant utility function; (ii) we calibrate the wage distribution to Swedish register data using the population distribution. We first calculate the welfare gains achievable by PP and tagging when the income tax is either linear or piecewise-linear. In these settings the welfare gains are calculated using as a benchmark equilibrium the allocation that one obtains under an optimal linear income tax. Second, we calculate the welfare gains from PP and tagging using an extended version of the general income tax model studied in section 2. In this case the welfare gains are calculated using as a benchmark the allocation obtained under an optimal nonlinear income tax. The reason for considering various types of income tax schedules is that the gains from PP (or tagging) might be overstated in the discrete type fully nonlinear tax model, given that it empowers the government with a (unrealistically) great deal of sophistication in the design of the tax schedule. Regarding the social welfare function, we focus on the max-min.[26]

4.1. Data

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 day 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. According to this definition in 2005 the fraction of parents in Sweden was slightly below 10%.[27] As an estimate of the hourly price for child care we have chosen a price of 40% of the median wage for parents. Since we have shown that the distribution of users and non-users matters for the welfare gains

[26]The max-min social welfare criteria avoids confounding the curvature of the individual utility function and preferences for redistribution and allows for the cleanest possible measure of the welfare gain of policy reform. This social welfare function is widely applied in the theoretical optimal taxation literature.

[27]Data has been combined from three sources, "Flergenerationsregistret", "Louise-databasen" and "Lonestrukturstatistiken". These statistics cover 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.

achievable through either PP or tagging, we also report results for the case where the proportion of parents is increased to 15%.[28]

4.2. Preferences

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:[29]

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

where $\alpha, \beta < 0$, $\gamma, \delta, \epsilon > 0$.[30] The annual time endowment J is set to 5840 hours. The labor supply function is:

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

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

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

We consider the parametrization given by $\alpha = -1.1$, $\beta = -0.0095$, $\gamma = 0.07$, $\delta = 0.95$ and $\epsilon = 2000$.[31]

[28] A possible interpretation of this exercise is the following. According to Blomquist et al. (2010) child care services and elderly care services represent the best examples of private goods fitting their model of PP. Applying this idea to our model, the group of users could be thought as being composed of people with small children and of people with elderly relatives who need to be taken care of. Thus, increasing the fraction of users from 8 to 15% might be interpreted as a way to measure the welfare gains achievable by publicly providing both child care- and elderly-care services. Admittedly, the measure that we get represents only a crude estimate of the welfare effects. The reason is that it rests on two implicit assumptions that are unlikely to be satisfied in practice. The first is that the unitary price of child care services and the unitary price of elderly care services are the same. The second is that users either need child-care or elderly-care services but not both at the same time. Notice however that, once public provision of child care services is supplemented by public provision of elderly care services, the relative merits of PP, as compared to tagging, are likely to be magnified. The reason is that if one can in principle think at the implementation of a tagging scheme that offers different tax schedules to parents and non-parents, it seems unfeasible to implement a tagging scheme that discriminates between agents who have to take care of their older relatives and agents who do not.

[29] A similar utility function is described by Stern (1986) as a good candidate for representing labor supply behavior. The quadratic specification has also recently 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.

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

[31] The parameterization is chosen with several features of labor supply behavior in mind. In particular we want the model to produce empirally relevant substitution- and income effects. We also want the labor supply curve to imply that individuals work a reasonable fraction of their time endowment.

The resulting uncompensated labor supply elasticity as a function of the wage rate is shown in Figure 1. Given that the distribution of wages for parents lies to the left of the wage distribution for non-parents, and that in the model parents are 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.[32] The income elasticities of labor supply are shown in Figure 2. Finally, in Figure 3 the labor supply function is graphed.[33] 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.[34]

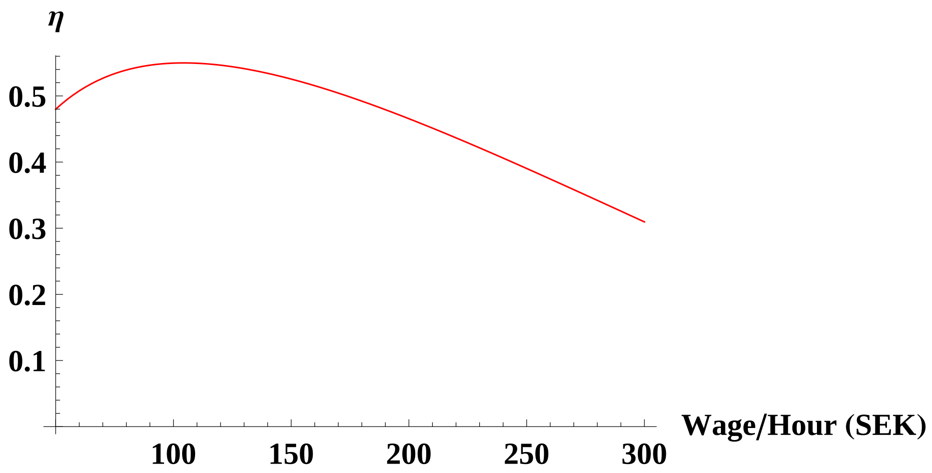

Figure 1: The uncompensated labor supply elasticity η over the support of the wage distribution.

4.3. Linear and piece-wise linear taxation

To save space we only present here the government's problem when the piece-wise linear tax is supplemented by a PP scheme. In appendix B we provide a more detailed

[32]See e.g. the review of the literature provided by Meghir and Phillips (2010).

[33]The labor supply function is evaluated at (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 problem.

[34]One 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 compensated labor supply elasticity generated by our utility function is decreasing in the wage rate of agents.

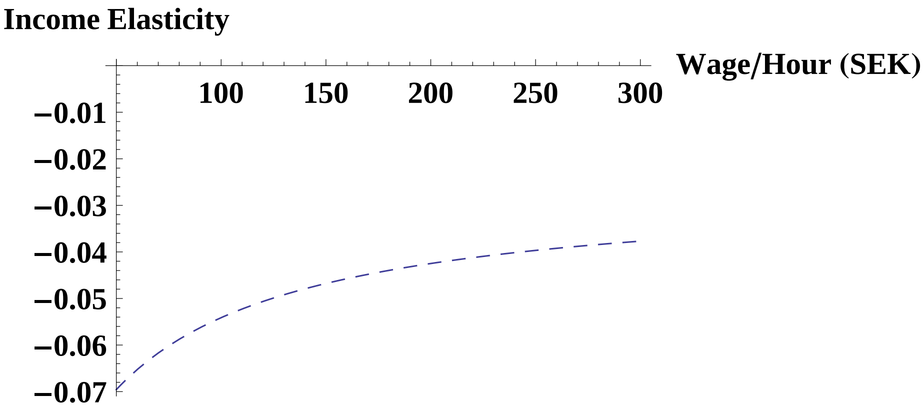

Figure 2: The income elasticity of labor supply over the support of the wage distribution.

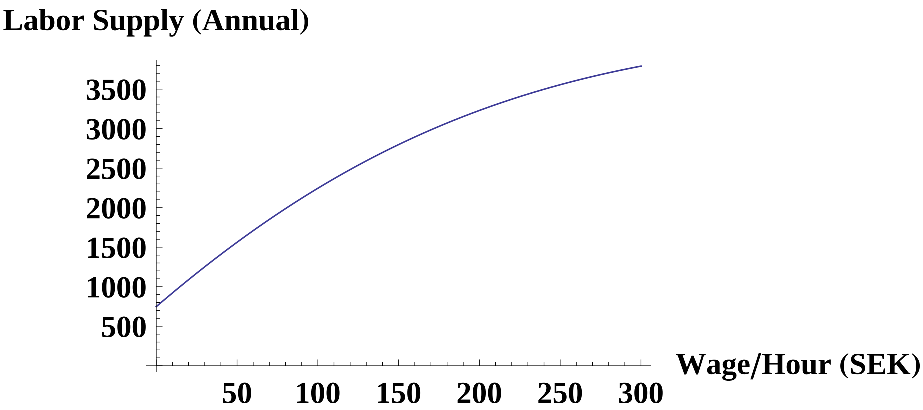

Figure 3: The labor supply function as a function of hourly wages.

characterization of PP under a linear income tax.[35] We approximate the max-min objective with the maximization of the demogrant. This is always valid when the least well-off individual does not work. In the case of tagging, the government designs two separate income tax schedules for the parent and non-parent groups respectively and can transfer resources across the groups; hence the max-min objective implies for tagging that the utility of the least well-off individual has to be the same in each group. When these agents do not work this means that the demogrant is the same for both groups.

As in section 2, let's assume that, in terms of ability types, there are in the population

[35] Always for space constraints we skip the formal characterization of the government's problem when tagging supplements a linear- or a piecewise-linear income tax.

n different types of agents with $w^1 < w^2 < ... < w^n$.[36] 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. We allow for a demogrant which is denoted G. This tax function can be described by the set of parameters

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

Under PP all agents, irrespective of whether they are parents or not, face the same budget constraint. For any $\theta \in \Theta_{pw}$ the budget constraint of an individual is $C(Y) = Y - T(Y;\theta)$, with $T(Y;\theta)$ defined as:

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

Given that $Y = wh$, agents choose h to maximize $U(C(Y), Y)$ and this leads to the following indirect utility function:

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

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

$$\max_{\theta \in \Theta} W(\theta) = \max_{\theta \in \Theta} V\left(\theta, w^1\right), \tag{8}$$

subject to the resource constraint:

$$\sum_{i=1}^{n} \pi^i \left(Y^{i*}(\theta; w^i) - C^{i*}(\theta; w^i)\right) \geq q \sum_{i=1}^{n} \pi^i \delta^i \frac{Y^{i*}(\theta; w^i)}{w^i},$$

where the proportion of agents of ability type i is π^i and $\delta^i \in [0, 1]$ denotes the proportion of parents among agents of ability type i.

[36] In the simulation we approximate the actual wage distributions in the data with $n = 1000$ taxpayer types.

The solution to the problem above yields the optimal piece-wise linear tax schedule $T^* = T(Y; \theta^*)$, where θ^* solves (8).[37] We solve this problem using numerical optimization techniques[38]. In order to get a tractable computational exercise we do not optimize over the break-points but instead take these as exogenously given. The break points are found by first solving the linear tax problem and then dividing the resulting income distribution into four brackets so that, in the linear tax optimum, an equal number of agents report income within each bracket.[39]

4.3.1. Results

Table 4.1: Linear and Piece-wise Linear Results

	t_1	t_2	t_3	t_4	G	Welfare Gain
Linear OT	64.97%	–	–	–	154849	–
Linear PP	67.51%	–	–	–	154984	0.06%
Linear Tagging (p)	46.71%	–	–	–	155933	
Linear Tagging (np)	66.38%	–	–	–	155933	0.49%
Piece-wise OT	99.39%	-1.31%	47.93%	41.33%	191715	16.69%
Piece-wise PP	99.54%	-2.20%	48.95%	42.51%	196317	18.77%
Piece-wise TG (p)	75.00%	-15.13%	33.68%	31.16%	196452	
Piece-wise TG (np)	99.96%	1.75%	49.38%	41.61%	196452	18.83%

The results are reported in Tables 4.1 and 4.2. Regarding the marginal tax rates we can notice that, under a piecewise linear income tax, they are neither always increasing nor always decreasing. In all cases they abide by the following pattern: starting with high marginal tax rates on the first bracket, they decrease substantially on the second bracket and from there they follow an inverted U-shaped profile. This is illustrated in Figure 4 for the OT case.

[37] This is not a concave programming problem. Although utility is continuous in Θ_{pw}, 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 was used.

[38] We represent the population distribution with 1998 agents represented by 999 wage rates from each group. These correspond to the quantiles of each distribution where we have excluded the extreme values of each distribution.

[39] If the (discrete) income distribution is represented by the vector y and the economy is populated by 100 agents, then the break-points chosen would be y_{25}, y_{50} and y_{75}. Agents are of course free to choose in which bracket they wish to locate.

Table 4.2: Welfare comparisons with 15% parents

Optimum	Welfare Gain
Linear PP	0.09%
Linear TG	0.84%
Piece-wise OT	15.26%
Piece-wise PP	19.01%
Piece-wise TG	18.97%

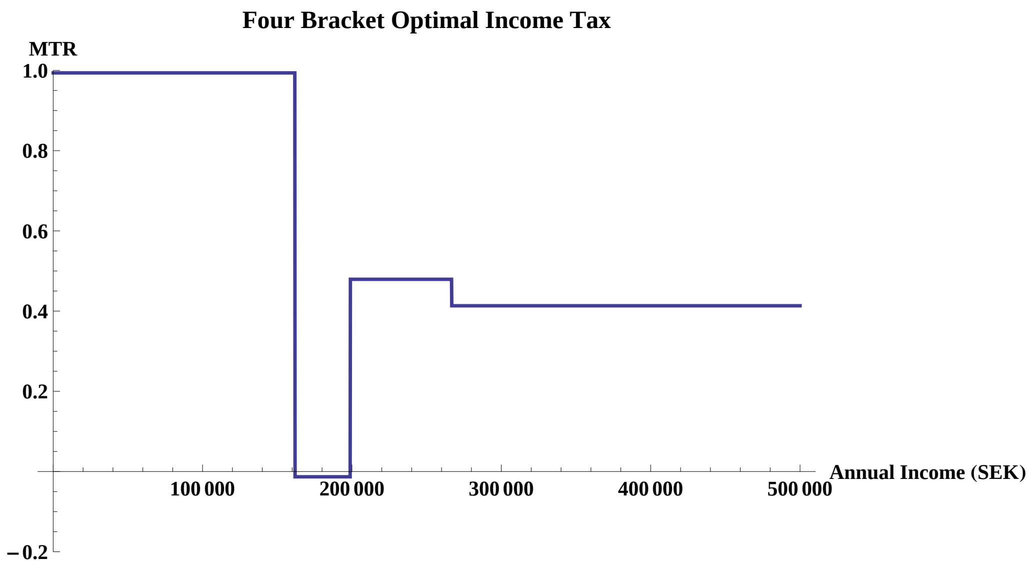

Figure 4: Optimal marginal tax rates for the four segment tax.

We can also see that both PP and tagging are always welfare-improving, even though the welfare gain from PP is very small when the income tax is linear. This should not be surprising and in fact with different parameter values one might have even obtained that PP is welfare-reducing when the government is restricted to a linear income tax. The reason is that the PP scheme delivers *per se* large welfare gains to all parents, despite the fact that the government is in our example only interested in maximizing the well-being of the least skilled among parents. Under a fully nonlinear income tax the government is able to adjust the income tax schedule in such a way to offset the welfare gain granted by PP to the non least skilled parents, while at the same time reaping the benefits delivered by PP in terms of slackened self-selection constraints. On the contrary, under a linear income tax the government lacks this flexibility in adjusting the income tax schedule. When a piecewise linear income tax is used, the government recovers at least part of the flexibility. Accordingly, moving from a linear income tax to a 4-brackets piecewise linear income tax, the difference between the welfare gains achievable through PP and those achievable through tagging tends to become virtually negligible.

Comparing the results presented in tables 4.1 and 4.2, we can also see that an hypothetical increase in the proportion of parents strengthens to a similar extent the welfare-enhancing power of PP and tagging. More precisely, a doubling in the proportion of parents would generate welfare gains which are, both for PP and for tagging, almost twice as large than in the baseline scenario.

Finally, Tables 4.1 and 4.2 show that, albeit the welfare gains from PP or tagging are substantial, much larger welfare gains can be reaped by switching from a linear- to a 4-brackets piecewise-linear income tax.

4.4. General (fully nonlinear) income tax

We extend the model from section 2 by allowing for six types and calibrate these wages to the data.[40] We also employ the utility function presented in section 4.2.

The (hourly) wages are given by:

$$w = (104.41 \quad 110.29 \quad 117.71 \quad 126.85 \quad 138.57 \quad 154.11),$$

where each ability type is represented by a fraction π_i of the population with:

$$\pi = (0.0284 \quad 0.3049 \quad 0.0284 \quad 0.3049 \quad 0.0284 \quad 0.3049).$$

Finally, letting $\Delta_i = 1$ indicate that type i is a parent, the pattern of parents which according to our definition arises from the (actual) economy is:[41]

$$\Delta = (1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 0).$$

[40]The wage distributions for parents and non-parents are each approximated by three wage levels corresponding to the 25:th, 50:th and 75:th percentiles of the population distribution. The reason for not considering more than 6 types is computational, as we cannot rely on the single crossing condition in the optimal taxation case without PP. We still believe that six types makes possible a fairly accurate approximation of the wage distributions.

[41]This pattern resembles more closely the "Case 2" from the theoretical part of the paper.

4.4.1. Results

The results for this case and a max-min social welfare function are summarized in Tables 4.3 and 4.4 below. After having observed that the welfare gains from both tagging and PP are increasing in the proportion of parents, the main insight from Table 4.4 is that the welfare gains from the two schemes are almost identical under a max-min social welfare function.

Table 4.3: General Nonlinear Tax - Calibrated 8% Parents

		p	np	p	np	p	np
	B_1	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
OT	199560	0	216240	0	297300	0	508740
PP	200220	70260	213300	213300	295860	295860	507720
TG	266193	129420	215580	185340	296520	326700	507540
		$T'(Y^1)$	$T'(Y^2)$	$T'(Y^3)$	$T'(Y^4)$	$T'(Y^5)$	$T'(Y^6)$
OT		88.12%	39.94%	89.46%	29.45%	91.05%	0%
PP		100%	41.01%	50.34%	29.73%	43.63%	0%
dist. part		*54.91%*	*41.01%*	*10.34%*	*29.73%*	*9.65%*	*0%*
TG		29.01%	39.91%	21.44%	29.43%	0%	0%

Table 4.4: Welfare Gains for the General Nonlinear Tax

	Proportion of Parents	
	8%	15%
Public Provision	2.55%	4.84%
Tagging	2.69%	5.11%

5. Concluding remarks

Previous literature has shown that, in the presence of a nonlinear income tax, public provision of 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 extended earlier analyses in three important respects. First, earlier studies only considered models with uni-dimensional heterogeneity. In this paper, focusing on child care services as an example of private good candidate for public provision, we have allowed for bi-dimensional heterogeneity, and have assumed that individuals vary both in terms of productivity and in the need for the publicly provided good. Second, we have assessed the magnitude of the welfare gains that can be achieved by public provision of private goods. Third, we have compared the welfare gains from public provision with the welfare gains achievable through tagging.

To accomplish the first task we have set up a theoretical model with both people who need child care, parents, and people who do not need child care, non-parents. In such a model several types of mimicking behaviors are possible: parents mimicking other parents, non-parents mimicking parents, non-parents mimicking other non-parents, and parents mimicking non-parents. When public provision is marginally introduced, financed by increased taxes, the two first types of self-selection constraints are mitigated. The third type is unaffected while the fourth is reinforced. Therefore, when we have both a group that needs the publicly provided good and another that does not, it is not obvious that public provision is welfare-enhancing. Whether public provision is welfare-enhancing or not depends to a large extent on how the wage distributions for users and non-users overlap.

A second contribution is that we are the first to quantitatively assess the welfare gains of publicly provided child care. This has been done using two different types of models. One highly stylized model has been used to study the sensitivity of results on the degree of overlap between the wage distributions for parents and non-parents and on the degree of social inequality to aversion. We have shown that the welfare gains are largest when three conditions are jointly met: first, there is a large degree of overlap between the wage distributions for parents and non-parents; second, parents are on average less productive than non-parents; third, the social aversion to inequality is high. We have also used a realistically calibrated model, based on the Swedish wage distribution, to assess whether the child care provision is welfare-improving or not. We have done this for tax systems of varying sophistication. We have found that the welfare gains are positive but almost nil when the income tax at disposal of the government is linear. If however we add public provision to an optimal nonlinear income tax the welfare gains are about 2.6% of aggregate consumption. Real income tax systems are more sophisticated than a linear income tax, but not as sophisticated as an optimal nonlinear income tax. We have therefore also studied the gains of public provision under an optimal piece-wise linear tax with 4 segments. The gains are then about 2%.

A third contribution is that we have compared the gains of public provision with the gains from tagging. We find this comparison particularly interesting since tagging schemes and public provision schemes are likely to have very different political appeal. For instance, if one considers child care services as a candidate for public provision, the tagging counterpart would be a scheme that assigns different tax schedules to women with children in day care ages; not exactly a gender based tax but a variant on the same theme. Thus, one of the purposes in comparing the effects of public provision and tagging was to investigate whether public provision could serve as a reasonably good, but politically more palatable, substitute for tagging schemes. We have found that, even if public provision seldom outperforms tagging, the difference in welfare gains tends to shrink the less the government is constrained in the design of the income tax schedule. Results from a model calibrated on Swedish data show that, with a general (unrestricted nonlinear) income tax and a max-min social welfare function, the welfare gains of the two schemes are almost identical and close to 2.6% of aggregate consumption.

Finally, our analysis has indicated that, albeit the welfare gains from public provision or tagging are substantial, they are small compared to the gains that can be reaped by switching from a linear- to a nonlinear income tax (about one sixth).

[1] 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.
[2] Balestrino, A. (2000), "Mixed tax systems and the public provision of private goods", International Tax and Public Finance 7, 463-478.
[3] Blomquist, S. and V. Christiansen (1995), "Public provision of private goods as a redistributive device in an optimum income tax model", Scandinavian Journal of Economics 97, 547-567.
[4] Blomquist, S. and V. Christiansen (2008), "Taxation and heterogenous preferences", FinanzArchiv 64, 218-244.
[5] Blomquist, S., V. Christiansen and L. Micheletto (2010), "Public provision of private goods and nondistortionary marginal tax rates", American Economic Journal: Economic Policy 2 (2), 1-27.
[6] Boadway, R. and M. Marchand (1995), "The use of public expenditures for redistributive purposes", Oxford Economic Papers 47, 45-59.
[7] Boadway R., M. Marchand, P. Pestieau and M. del Mar Racionero (2002), "Optimal redistribution with heterogeneous preferences for leisure", Journal of Public Economic Theory 4, 475-498.
[8] Boadway, R. and P. Pestieau (2006), "Tagging and redistributive taxation", Annales d'Economie et de Statistique 83/84, 123-147.
[9] Cremer, H. and F. Gahvari (1997), "In-kind transfers, self-selection and optimal tax policy", European Economic Review 41, 97-114.
[10] Cremer, H., F. Gahvari and J. M. Lozachmeur (2010), "Tagging and income taxation: theory and an application", American Economic Journal: Economic Policy 2 (1), 31-50.
[11] Currie, J. and F. Gahvari (2008), "Transfers in cash and in-kind: theory meets the data", Journal of Economic Literature 46, 333-383.
[12] Guesnerie, R. and K. Roberts (1984), "Effective policy tools and quantity controls", Econometrica 52, 59-82.
[13] Guesnerie, R. and J. Seade (1982), "Nonlinear pricing in a finite economy", Journal of Public Economics 17, 157-179.
[14] Immonen, R., R. Kanbur, M. Keen and M. Tuomala (1998), "Tagging and taxing: The use of categorical and income information in designing tax/transfer schemes", Economica 65, 179-192.
[15] Jordahl, H. and L. Micheletto (2005), "Optimal utilitarian taxation and horizontal equity", Journal of Public Economic Theory 7, 681-708.
[16] Kaplow, L. (2008), "Optimal policy with heterogeneous preferences", B.E. Journal of Economic Analysis and Policy: Advances, Issue 1, Article 40, 1-28.
[17] Meghir, C. and D. Phillips (2010), "Labour supply and taxes", in Dimensions of Tax Design: The Mirrlees Review, OUP Oxford.
[18] Micheletto, L. (2004), "Optimal redistributive policy with endogenous wages", FinanzArchiv 60, 141-159.
[19] Mirrlees, J. A. (1971), "An exploration in the theory of optimum income taxation", Review of Economic Studies 38, 175-208.
[20] Nichols, A. L. and R. J. Zeckhauser (1982), "Targeting transfers through restrictions on recipients", American Economic Review 72, Papers and Proceedings, 372-377.
[21] Pirttil, J. and M. Tuomala (2002), "Publicly provided private goods and redistribution: a general equilibrium analysis", Scandinavian Journal of Economics 104, 173-188.
[22] Saez, E. (2002), "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes", Journal of Public Economics 83, 217-230.
[23] Sandmo, A. (1993), "Optimal redistribution when tastes differ", FinanzArchiv 50, 149-163.
[24] Sheshinski, E. (1972), "The optimal linear income tax", Review of Economic Studies 39, 297-302.
[25] Slemrod, J., S. Yitzhaki, J. Mayshar and M. Lundholm (1994), "The optimal two-bracket linear income tax", Journal of Public Economics 53, 269-290.
[26] Stern, N. H. (1976), "On the specification of models of optimum income taxation", Journal of Public Economics 6, 123-162.
[27] Stern, N. H. (1982), "Optimum taxation with errors in administration", Journal of Public Economics 17, 181-211.
[28] Stern, N. H. (1986), "On the specification of labour supply functions", in *Unemployment, Search and Labour Supply*, ed. R. W. Blundell and I. Walker, 121-142. Cambridge: Cambridge University Press.
[29] Stiglitz, J. E. (1982), "Self-selection and Pareto-efficient taxation", Journal of Public Economics 17, 213-240.
[30] Tarkiainen, R. and M. Tuomala, (1999), "Optimal nonlinear income taxation with a two-dimensional population; A computational approach", Computational Economics 13, 1-16.

[31] Tuomala, M. (2010), "On optimal non-linear income taxation: numerical results revisited", International Tax and Public Finance 17, 259-270.

Appendix A. Numerical results

In this appendix the numerical results for the discrete model of section 2 are presented. The problem was set up in AMPL and then solved using the KNITRO package for constrained optimization developed by Ziena Optimization Inc. Because the problem is of the multidimensional heterogeneity type the full set of incentive constraints has been included in the optimization.

Table A.1: Allocations Nonlinear Taxation (CASE 1, Utilitarian)

		Proportion of Parents			
		0.15		0.30	
Type		$T'(Y)$	Y	$T'(Y)$	Y
1 (Parent)	OT	22.17 %	0.460	20.42 %	0.507
	PP	62.36 % (12.36%)	0.504	61.60 % (11.60%)	0.537
	TG	2.14 %	0.468	2.50 %	0.494
	TG+PP	51.05 %	0.469	51.23 %	0.496
2 (Parent)	OT	53.63 %	0.460	51.23 %	0.540
	PP	61.11 % (25.39%)	0.827	59.37 % (23.66%)	0.885
	TG	0.00 %	0.856	-0.00 %	0.900
	TG+PP	35.71 %	0.863	35.71 %	0.908
3 (Nonparent)	OT	14.32 %	1.977	19.82 %	1.989
	PP	13.19 %	1.990	17.48 %	2.022
	TG	10.20 %	2.061	11.61 %	2.183
	TG+PP	10.20 %	2.061	11.60 %	2.183
4 (Nonparent)	OT	-0.00 %	3.152	-0.00 %	3.233
	PP	0.00 %	3.154	-0.00 %	3.241
	TG	0.00 %	3.202	0.00 %	3.352
	TG+PP	0.00 %	3.201	0.00 %	3.351
Welfare Gains					
R_{PP}		1.04%		2.38%	
R_{TG}		2.27%		4.99%	
R_{PPTG}		2.27%		5.00%	

Table A.2: Allocations Nonlinear Taxation (CASE 2, Utilitarian)

		Proportion of Parents			
		0.15		0.30	
Type		$T'(Y)$	Y	$T'(Y)$	Y
1 (Parent)	OT	36.60 %	0.334	36.50 %	0.356
	PP	61.45 % (11.45%)	0.533	61.45 % (11.45%)	0.559
	TG	8.91 %	0.481	9.88 %	0.499
	TG+PP	54.45 %	0.499	54.94 %	0.519
2 (Parent)	OT	54.90 %	1.192	44.24 %	1.493
	PP	64.79 % (40.98%)	1.389	52.16 % (28.35%)	1.641
	TG	0.00 %	1.753	-0.00 %	1.812
	TG+PP	23.81 %	1.779	23.81 %	1.842
3 (Nonparent)	OT	13.45 %	1.192	13.99 %	1.233
	PP	17.03 %	1.182	21.00 %	1.204
	TG	19.45 %	1.176	21.36 %	1.226
	TG+PP	19.42 %	1.176	21.28 %	1.224
4 (Nonparent)	OT	0.00 %	3.271	0.00 %	3.285
	PP	-0.00 %	3.273	-0.00 %	3.320
	TG	0.00 %	3.316	0.00 %	3.424
	TG+PP	0.00 %	3.314	0.00 %	3.420
Welfare Gains					
R_{PP}		0.89%		2.14%	
R_{TG}		1.96%		3.99%	
R_{PPTG}		2.02%		4.15%	

Table A.3: Allocations Nonlinear Taxation (CASE 3, Utilitarian)

		Proportion of Parents			
		0.15		0.30	
Type		$T'(Y)$	Y	$T'(Y)$	Y
1 (Parent)	OT	35.34 %	0.180	32.91 %	0.202
	PP	64.69 % (14.69%)	0.240	64.00 % (14.00%)	0.253
	TG	6.86 %	0.229	7.73 %	0.239
	TG+PP	53.41 %	0.235	53.85 %	0.246
2 (Parent)	OT	9.92 %	0.913	13.14 %	0.915
	PP	15.28 % (-9.72%)	0.925	19.26 % (-5.74%)	0.936
	TG	0.00 %	0.772	-0.00 %	0.803
	TG+PP	25.00 %	0.781	25.00 %	0.813
3 (Nonparent)	OT	15.91 %	0.913	18.99 %	0.915
	PP	15.28 %	0.925	19.26 %	0.936
	TG	11.61 %	0.960	12.94 %	1.006
	TG+PP	11.60 %	0.959	12.91 %	1.005
4 (Nonparent)	OT	0.00 %	1.569	0.00 %	1.590
	PP	0.00 %	1.581	-0.00 %	1.617
	TG	0.00 %	1.600	-0.00 %	1.662
	TG+PP	0.00 %	1.599	0.00 %	1.660
Welfare Gains					
R_{PP}		1.47%		2.88%	
R_{TG}		2.15%		4.18%	
R_{PPTG}		2.18%		4.28%	

Table A.4: Allocations Nonlinear Taxation (CASE 1, Max-min)

		Proportion of Parents			
		0.15		0.30	
Type		$T'(Y)$	Y	$T'(Y)$	Y
1 (Parent)	OT	50.00 %	0.000	50.00 %	0.000
	PP	95.81 % (45.81%)	0.166	91.77 % (41.77%)	0.246
	TG	29.94 %	0.357	30.20 %	0.376
	TG+PP	70.76 %	0.429	70.95 %	0.451
2 (Parent)	OT	64.29 %	0.000	64.29 %	0.000
	PP	94.80 % (59.08%)	0.305	89.02 % (53.31%)	0.469
	TG	0.00 %	0.987	0.00 %	1.033
	TG+PP	35.71 %	0.996	35.71 %	1.043
3 (Nonparent)	OT	48.68 %	1.635	49.74 %	1.717
	PP	48.61 %	1.630	49.47 %	1.696
	TG	48.55 %	1.625	49.40 %	1.691
	TG+PP	48.53 %	1.623	49.36 %	1.688
4 (Nonparent)	OT	-0.00 %	3.244	-0.00 %	3.369
	PP	-0.00 %	3.235	-0.00 %	3.338
	TG	-0.00 %	3.227	0.00 %	3.329
	TG+PP	-0.00 %	3.225	0.00 %	3.325
Welfare Gains					
R_{PP}		0.68%		2.39%	
R_{TG}		1.90%		4.65%	
R_{PPTG}		2.13%		5.20%	

Table A.5: Allocations Nonlinear Taxation (CASE 2, Max-min)

		Proportion of Parents			
		0.15		0.30	
Type		$T'(Y)$	Y	$T'(Y)$	Y
1 (Parent)	OT	50.00 %	0.000	50.00 %	0.000
	PP	95.78 % (45.78%)	0.175	91.73 % (41.73%)	0.257
	TG	42.49 %	0.231	42.95 %	0.233
	TG+PP	76.84 %	0.401	77.12 %	0.414
2 (Parent)	OT	49.07 %	1.335	49.39 %	1.384
	PP	87.08 % (63.27%)	0.892	78.10 % (54.29%)	1.199
	TG	-0.00 %	1.927	0.00 %	1.983
	TG+PP	23.81 %	1.943	23.81 %	1.997
3 (Nonparent)	OT	0.00 %	1.281	0.00 %	1.325
	PP	56.41 %	0.892	59.88 %	0.899
	TG	58.69 %	0.864	59.46 %	0.890
	TG+PP	58.65 %	0.863	59.37 %	0.887
4 (Nonparent)	OT	0.00 %	3.247	-0.00 %	3.311
	PP	0.00 %	3.366	0.00 %	3.439
	TG	0.00 %	3.355	0.00 %	3.432
	TG+PP	-0.00 %	3.351	0.00 %	3.423
Welfare Gains					
R_{PP}		2.52%		4.63%	
R_{TG}		4.47%		7.96%	
R_{PPTG}		4.93%		8.89%	

Table A.6: Allocations Nonlinear Taxation (CASE 3, Max-min)

		Proportion of Parents			
		0.15		0.30	
Type		$T'(Y)$	Y	$T'(Y)$	Y
1 (Parent)	OT	50.00 %	0.000	50.00 %	0.000
	PP	96.89 % (46.89%)	0.071	93.79 % (43.79%)	0.105
	TG	40.88 %	0.121	41.33 %	0.124
	TG+PP	76.08 %	0.195	76.36 %	0.202
2 (Parent)	OT	75.00 %	0.000	75.00 %	0.000
	PP	48.47 % (23.47%)	0.760	46.55 % (21.55%)	0.800
	TG	0.00 %	0.866	0.00 %	0.896
	TG+PP	25.00 %	0.874	25.00 %	0.903
3 (Nonparent)	OT	50.89 %	0.756	52.02 %	0.793
	PP	48.47 %	0.760	46.55 %	0.800
	TG	50.63 %	0.747	51.36 %	0.772
	TG+PP	50.60 %	0.746	51.29 %	0.769
4 (Nonparent)	OT	-0.00 %	1.633	0.00 %	1.696
	PP	-0.00 %	1.614	-0.00 %	1.651
	TG	-0.00 %	1.617	0.00 %	1.659
	TG+PP	-0.00 %	1.616	0.00 %	1.656
Welfare Gains					
R_{PP}		3.48%		8.64%	
R_{TG}		3.46%		8.54%	
R_{PPTG}		3.90%		9.45%	

Appendix B. Characterization of public provision under a linear income tax

Also for the case of a linear income tax we limit ourselves to the analysis of the government's problem in the presence of public provision. We then resort to numerical simulations in order to get a quantitative assessment of the potential welfare-gains descending from the implementation of the public provision scheme. Before proceeding, however, it will be useful to make a remark about the desirability of a public provision scheme in a model of linear income taxation. In the first part of this paper we have discussed how the public provision scheme that we analyze might be welfare-enhancing in the presence of a nonlinear income tax due to the effects that it implies on the binding self-selection constraints. One might then wonder if our public provision scheme still retains the potential of being welfare-improving in a setting where income is taxed on a linear scale and therefore no self-selection constraint explicitly appears in the formulation of the government's problem. The answer would be clearly negative in a setting where all agents were parents, and therefore users of the publicly provided service. However, in a setting with both parents and non-parents, the public provision scheme might still be welfare-enhancing since, as compared with a demogrant, it represents an instrument which only benefits parents. Thus, if the government places a sufficiently large weight on the welfare of parents, the public provision scheme might be desirable even in the presence of a linear income tax.

Under a linear income tax cum public provision, agents solve the problem $\max_h u(G + (1-t)wh, h)$, where G represents the uniform lump-sum transfer paid to all agents and t is the constant marginal income tax rate. Denoting by $V^i(t, G)$ the indirect utility of agents of ability type i, the design problem solved by the government can be written as:[42]

$$\max_{t,G} \quad \sum_{i=1}^{n} \beta^i V^i(t, G)$$

subject to:

[42]The indirect utility $V^i(t, G)$ is the solution to the problem $\max_h u\left(G + (1-t)w^i h, h\right)$.

$$t\sum_{i=1}^{n}\pi^i Y^i \geq G + q\sum_{i=1}^{n}\pi^i\delta^i Y^i/w^i, \qquad (\mu)$$

where μ is the Lagrange multiplier associated with the government's budget constraint.

The first order condition with respect to G is the following:

$$\sum_{i=1}^{n}\frac{\beta^i}{\mu}\frac{\partial V^i(t,G)}{\partial G} + \sum_{i=1}^{n}\left(t - q\frac{\delta^i}{w^i}\right)\pi^i\frac{\partial Y^i}{\partial G} = 1. \tag{B.1}$$

Defining by b^i the net social marginal valuation of a lump-sum transfer to an agent of ability type i, we have:

$$b^i \equiv \frac{1}{\pi^i}\frac{\beta^i}{\mu}\frac{\partial V^i(t,G)}{\partial G} + \left(t - q\frac{\delta^i}{w^i}\right)\frac{\partial Y^i}{\partial G}.$$

Having defined b^i we can easily see that condition (B.1) boils down to requiring $E(b) = 1$, where $E(\cdot)$ denotes the expectation operator.[43] In other words, it prescribes that at an optimum the lump-sum component should be adjusted such that b, the government's net social marginal valuation of a transfer of 1€ (measured in terms of government's revenue) should on average be equal to its marginal cost (1€).

The first order condition with respect to t is the following:

$$\sum_{i=1}^{n}\beta^i\frac{\partial V^i(t,G)}{\partial t} + \mu\left[\sum_{i=1}^{n}\pi^i Y^i + \sum_{i=1}^{n}\left(t - q\frac{\delta^i}{w^i}\right)\pi^i\frac{\partial Y^i}{\partial t}\right] = 0. \tag{B.2}$$

Using the Roy's identity $\frac{\partial V^i(t,G)}{\partial t} = -Y^i\frac{\partial V^i(t,G)}{\partial G}$ and the Slutsky equation $\frac{\partial h^i}{\partial t} = -w^i\frac{\partial h_s^i}{\partial(1-t)w^i} - Y^i\frac{\partial h^i}{\partial G}$, where the subscript s on h_s^i denotes a compensated response, after some rearrangements we can rewrite (B.2) as:

$$\sum_{i=1}^{n}\frac{\beta^i}{\mu}\frac{\partial V^i(t,G)}{\partial G}Y^i + \sum_{i=1}^{n}\left(t - q\frac{\delta^i}{w^i}\right)\pi^i\frac{\partial Y^i}{\partial G}Y^i - \sum_{i=1}^{n}\pi^i Y^i$$
$$= -\sum_{i=1}^{n}\left(t - q\frac{\delta^i}{w^i}\right)w^i\frac{\partial h_s^i}{\partial(1-t)w^i}\pi^i w^i.$$

[43] Apart 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.

Using the definition of b^i and remembering that at an optimum it must be that $E(b) = 1$, the condition above reduces to:

$$cov(b, Y) = -\sum_{i=1}^{n} \left(t - q\frac{\delta^i}{w^i}\right) w^i \frac{\partial h_s^i}{\partial (1-t) w^i} \pi^i w^i, \tag{B.3}$$

where $cov(\cdot, \cdot)$ denotes the covariance operator.

Denoting by $\overline{w}^i$ the marginal net wage rate of an agent of ability type i, $\overline{w}^i = (1-t) w^i$, and by $\epsilon^i_{h,\overline{w}^i}$ the compensated elasticity of labor, we can rewrite (B.3) as

$$-cov(b, Y) = \frac{1}{1-t} \sum_{i=1}^{n} \left(t - q\frac{\delta^i}{w^i}\right) \pi^i Y^i \epsilon^i_{h,\overline{w}^i}. \tag{B.4}$$

Notice that in the context of our model with public provision it is no longer possible to express the condition implicitly defining the optimal value for t in a way as simple as in a standard model without public provision (see e.g. Sheshinski (1972)). To express the condition for t in a way that mirrors as closely as possible the standard condition, we can rewrite (B.4) as:

$$\frac{t - \sum_{i=1}^{n} q^i \xi^i}{1-t} = -\frac{cov(b, Y)}{\sum_{i=1}^{n} \pi^i Y^i \epsilon^i_{h,\overline{w}^i}}, \tag{B.5}$$

where q^i has been defined as $q^i \equiv \delta^i q / w^i$ and ξ^i has been defined as $\xi^i \equiv \pi^i Y^i \epsilon^i_{h,\overline{w}^i} / \left(\sum_{i=1}^{n} \pi^i Y^i \epsilon^i_{h,\overline{w}^i} \right)$. Written in this form, we can see that the second term at the numerator of the left hand side of (B.5) is the new term that, due to the public provision scheme, enters the formula that implicitly defines the optimal value for t. In (B.5) ξ^i can be interpreted as the normalized earned income response of agents of ability type i to a marginal compensated increase in their after-tax wage rate.[44] As for q^i, it represents the marginal cost, in terms of higher (public) expenditures on publicly provided services, of an increase in the income earned by agents of ability type i. Since $\sum_{i=1}^{n} q^i \xi^i$ is unambiguously positive, the

[44] The reason why we use the word "normalized" is that at the denominator of the expression defining ξ^i we have the sum of the earned income responses of all agents to a marginal increase in their after-tax wage rate.

new term appearing in (B.5) tends to raise the optimal value of t for any given value of the right side of (B.5).

In our view, however, expressing the condition defining the optimal value for t in terms of the compensated elasticity of labor, as it is usually done, is in a setting with public provision not very informative.

A better alternative is to start with (B.3), recognize that $w^i \frac{\partial h_s^i}{\partial (1-t) w^i} = -\frac{\partial h_s^i}{\partial t}$, and then, defining φ^i as $\varphi^i \equiv \pi^i w^i \frac{\partial h_s^i}{\partial t} / \left(\sum_{i=1}^{n} \pi^i w^i \frac{\partial h_s^i}{\partial t} \right)$, express the condition defining the optimal t as:

$$t = \frac{cov\,(b, Y)}{\sum_{i=1}^{n} \pi^i w^i \frac{\partial h_s^i}{\partial t}} + \sum_{i=1}^{n} q^i \varphi^i. \tag{B.6}$$

In (B.6) φ^i can be interpreted as the normalized earned income response of agents of ability type i to a marginal compensated increase in the income tax rate. The virtue of (B.6) is that it defines in a simple way the optimal value of t as the sum of a term trading off equity and efficiency considerations (the first term on the right side of (B.6)) plus a corrective term given by a weighted average of the q^i-marginal costs.

Notice that we have already encountered the q^i-marginal costs in section 2.2, when we characterized the marginal tax rates prevailing at a nonlinear income tax optimum cum public provision. There, as one can see from (6)-(7), q^i (which remember has been defined as $\delta^i q / w^i$) only entered the expression characterizing the marginal tax rate faced by agents of ability type i. This happened because a nonlinear income tax allowed the government to let agents of different ability type face different marginal tax rates. Given that a linear income tax restricts the government to let all agents face the same marginal income tax rate, the term $\sum_{i=1}^{n} q^i \varphi^i$ can be regarded as the natural counterpart of the $\delta^i q / w^i$ appearing in (6)-(7).