Intergenerational transmission of skills during childhood and optimal fiscal policy*

Alessandra Casarico† Luca Micheletto‡ Alessandro Sommacal§

This draft: 26 April 2010

Abstract

The paper aims at characterizing the optimal tax policy and the optimal level of quality of day care in a two-type OLG model where parental choices over child care (that is, parental time devoted to children and time spent in day care centers) determine the probability of having a high skill child in a type-specific way. Parents derive utility from their own consumption, leisure, time spent with their kids and from the kids' expected human capital (warm-glow component). We consider two different scenarios: first, one where the government can use linear taxation on labor income and a linear tax/subsidy on day care. Second, a set-up where the government can resort to nonlinear taxation of labor income and again a linear tax/subsidy on day care. In both cases we discuss the rules dictating the optimal choice of day care quality enforced by the government. With respect to previous contributions, optimal tax formulas incorporate two new sets of terms. The first depends on the extent to which the social welfare function reflects the warm-glow component of parental preferences. The second depends on the social marginal utility of turning an unskilled individual into a skilled one.

KEYWORDS: optimal taxation; child care; intergenerational transmission of skills

JEL Classification: H21

*Corresponding author:

†Department of Institutional Analysis and Public Management, and Econpubblica, Bocconi University. E-mail address: alessandra.casarico@unibocconi.it

‡Faculty of Law, University of Milan, and Econpubblica, Bocconi University. E-mail address: luca.micheletto@unibocconi.it

§Faculty of Economics, University of Verona, and Econpubblica, Bocconi University. E-mail address: alessandro.sommacal@unibocconi.it

1 Introduction

The paper aims at characterizing the optimal tax policy and the optimal level of quality of day care in an OLG model where parental choices over child care (that is, parental time devoted to children and time spent in day care centers) determine the probability of a child to have high market ability. More specifically, we assume that agents are heterogeneous in two dimensions: market ability and ability to raise children. By ability to raise children we mean the ability to transfer cognitive and non-cognitive skills which are valuable on the market, for a given amount of time spent with the children themselves. Both abilities can be either high or low. The distribution of market ability/human capital across individuals is endogenous, that is, it depends on parental choices over child care, while the distribution of the ability to raise children is assumed to be exogenous.

The role of child care for children's human capital acquisition has been widely studied in the psychology and sociology literature. Economists have more recently recognized the importance of child care on skills' acquisition. This is documented by two recent strands of the literature. The first one describes the individual's skill formation (see Cuhna et al. 2005 for a review) as a dynamic process, characterized by strong complementarities between early and late investments in human capital (Carneiro and Heckman 2003; Cunha and Heckman 2007; Carneiro, Meghir, and Parey 2007). As there are critical and sensitive periods for the development of both cognitive and non-cognitive abilities, later remediation for early deficits in the formation of some important abilities is difficult and costly. Carneiro and Heckman (2002, 2003), for instance, suggest that the most important factor explaining the positive relation between income and college enrolment in the US is not related to short term liquidity constraints that poor individuals may face, but to the fact that they lived in early environments which were unable to form the cognitive and non-cognitive abilities required for success in school. A second strand of the literature looks at the importance of parental time, and especially maternal time, vs. other types of child care in producing children abilities. The earlier contributions - as surveyed for instance by Ruhm (2004) - reached mixed conclusions. More recent contributions, see for instance Bernal (2008) and Bernal and Keane (2007, 2008) highlight that, on average, the substitution of maternal time with other child care sources produces negative and rather sizable effects on children skills. However, they also show that this result masks some differences across types of child care and maternal education: for instance, formal care (i.e. center-based care and preschool) may have positive effects on children of poorly educated mothers. This is also documented in Heckman and Masterov (2007) who review the evidence supporting the idea that high quality preschool centers available to disadvantaged children on a voluntary basis are highly effective in promoting achievement for disadvantaged children. The indications of this literature do therefore support the appropriateness of including child care in the skill formation process and of allowing for a type-specific impact of parental time and day care on the accumulation of human capital.

In our model parents derive utility from their own consumption, leisure, time spent with their kids and from the kids' expected human capital (warm-glow component). The warm-glow assumption is consistent with altruism à la Andreoni (1989) and it is often used in papers focusing on the intergenerational transmission of human capital.[1] We consider two different scenarios: first, one where the government can use linear taxation on labor income and a linear tax/subsidy on day care. This system is equivalent to one where the two goods, consumption and day-care, are linearly taxed according to differentiated tax rates. Second, a set-up where the government can resort to nonlinear taxation of labor income and again a linear tax/subsidy on day care. In this case, to sidestep the complexities associated with multi-dimensional screening, we simplify our set-up and assume that there is perfect correlation between the two types of ability. We therefore move from a four- to a two-type agent model. A high (low) market ability type will in this case coincide with a high (low) ability to raise children type. Under this circumstance we will consider as more relevant the situation where children of low skilled individuals benefit from day care (that is, day care increases their probability of becoming high skilled), while children of high skilled individuals will see their probability of being high skilled tomorrow unaffected or reduced by the substitution of parental time with day care. In both frameworks, we discuss the rules dictating the optimal choice of day care quality enforced by the government.

Admittedly, the inclusion of day care services into a second-best optimal taxation framework is not a novelty of our analysis (see e.g. Balestrino 2000, Blomquist, Christiansen, and Micheletto 2010 and Blomquist and Micheletto 2009). However, when the optimal taxation literature discusses day care services, it does so by treating them just as one prominent exam-

[1] We are not alone in adopting warm-glow preferences: many papers on the intergenerational transmission of human capital and wealth share this assumption (*inter alia*, see Glomm and Ravikumar 1992; Galor and Zeira 1993; Glomm and Kaganovich 2003, 2008; Cremer and Pestieau 2006a). Though the empirical investigation of motives for transfers is not conclusive, the warm-glow of giving seems to be important in motivating agents' actions towards others (see Schokkaert 2006 for an exhaustive survey).

ple, among possible others, of goods/services that are complements to labor. As such, it has been suggested that their consumption should be encouraged by the tax system, or that they should be publicly provided, in order to either mitigate the distortion against labor supply determined by income taxation or to soften self-selection constraints in models of nonlinear income taxation. This way of looking at day care services is however, in our opinion, limited. To view them simply as an example of a complementary to labor item in the agents' consumption bundle prevents from recognizing other important roles which day care can play and which can be relevant for policy conclusions, such as its input in human capital accumulation. Under this respect, our paper contributes to the existing optimal taxation literature by trying to incorporate into the model this previously neglected aspect of day care services.

With respect to previous contributions, we find that optimal tax formulas incorporate a new Pigouvian term which corrects for the intergenerational externality in human capital accumulation stemming from the assumption of warm-glow altruism. Indeed, the warm-glow assumption delivers an inefficiency in the human capital formation process as parents do not fully internalize the effects of their time devoted to child care on the utility of their offsprings. This inefficiency calls for policy correction. The direction of this correction should ideally be type-specific, as the productivity of parental time in producing market skills depend on the parent's type. When only linear instruments are available to the government, the tax rates applied to consumption and day-care need to be the same across skill types. For this reason, the new term in the optimal tax formulas has to average the adjustments ideally required to correct the behavior of the four types of agents. When also a nonlinear tax on labor income is at the government's disposal, while the tax on day care is still linear, the marginal effective tax rate on labor income faced by high-skilled agents[2] is positive, implying that the overall effect of the tax system is to induce high-skilled agents to reduce their labor supply in order to spend more time with their children. The marginal effective tax rate on low-skilled agents tends to be lower than the one we would find in a model where the negative externality of low-skilled parental time on their kids' human capital is absent. This is a way to induce low-skilled agents to work more and substitute consumption for parental time. These corrections are required in order to induce the current period adults to internalize the social welfare effect generated by the link between their time allocation decision and the proportion of high-skilled adults in the

[2]High skilled refers here both to market ability and to ability to raise children, given the assumption of perfect correlation between the two.

next period.

We also find that the so called "principle of targeting" fails to hold in our model. This principle states that a distortion is best addressed by the instrument that acts directly on the relevant margin. In our setting it is the tax imposed on the purchases of day care services that can be interpreted as the direct instrument to correct the agents' behavior. Therefore, according to the principle of targeting, only the tax formula for this policy instrument should be modified for corrective purposes. However, as we will see, all the tax formulas are affected. Intuitively, this happens because, whereas the required correction to the behavior of parents is type-specific, the tax rate on day care purchases is restricted to be linear and therefore uniform across types.

As far as the optimal choice of the quality of day care is concerned, the latter is determined by equating the total private marginal benefits of a quality increase to the marginal costs plus two additional terms: the first one captures the impact of a change in the quality on the government budget constraint through the change in demand of consumption and day care services. The second one reflects, as above, the intergenerational externality in human capital accumulation.

Both in the determination of the optimal tax formulas and of the optimal quality of day care we allow for a merit good term which accounts for the possibility that the government preferences deviate from the individual ones, that is, we allow for the government to disregard the warm-glow component of individual utility as it is done, for instance in Cremer and Pestieau (2006b). They analyze the optimal tax policy in a dynamic OLG model where the probability of a child to be skilled is affected by education expenditures of parents motivated by warm glow altruism. The crucial difference is that, in our framework, the way parents' choices affect the level of human capital of the respective offspring depends on the parents' skills in a type-specific way. As argued above, this assumption has important implications for the design of the optimal tax system.

The paper is organized as follows. In section 2 we present the basic ingredients of the model and we describe the behavior of agents, the productive technology, the evolution over time of the skill distribution in the population and the government's objective function. In section 3 we analyze the solution to the government's problem under a linear tax system. In section 4 we consider the possibility of a so-called mixed tax system where earned income can be subject to a nonlinear tax function whereas commodity purchases are restricted to be taxed according to a set of differentiated but linear commodity taxes. Finally, section 5 contains a discussion of some possible extensions of our model and section 6 offers concluding remarks.

2 The model

2.1 The consumers

We consider a two-period OLG model with bi-dimensional intragenerational heterogeneity: agents differ in their market ability, that is, in their human capital, and in their ability to raise children. By ability to raise children we mean the ability to transfer human capital for a given amount of time spent with the children. While the distribution of human capital is endogenous and it depends on child care arrangements in a way which will be specified further down, the distribution of the latter is assumed to be exogenous. In the first period agents (children) do not take any active choice; depending on child care arrangements, on the human capital of their parents and on the ability of parents to raise children, they have a certain probability to have a high or a low level of human capital. In the second period agents, given their level of skills, decide how to allocate their time between labor, time devoted to children and leisure. Each adult is assumed to have a child and parents of market ability type j and ability to raise children type k maximize the following utility function:

$$U_t^{jk} = u\left(c_t^{jk}, z_t^{jk}, n_t^{jk}\right) + \eta\left(\pi^{jk}(n_t^{jk})H^2 + (1 - \pi^{jk}(n_t^{jk}))H^1\right), \tag{1}$$

with $\eta''(\cdot) < 0 < \eta'(\cdot)$, $u''(\cdot) < 0 < u'(\cdot)$, and where c_t^{jk}, z_t^{jk}, n_t^{jk} denote respectively consumption, leisure and time devoted to child care by an agent jk. As regards market ability, there are only two possible levels of it and they are denoted by H^j, $j = 1, 2$, with $H^2 > H^1$. As regards the ability to raise children, it can only take two possible values, low and high, respectively denoted by $k = 1$ and $k = 2$.[3] Parents with high (resp.: low) ability to raise children have, other things being equal, a higher (resp.: lower) probability to raise a child who will become a high market ability adult. The last term in (1) reflects the warm-glow altruism of parents (Andreoni 1989), who care about the impact that their parental time will have on the probability π^{jk} of having a high-human capital child and therefore on the expected level of market ability of their kid. $\pi^{jk}(n_t^{jk})$ stands for $\pi^k(n_t^{jk}, H^j, e_t)$, that is the probability of

[3]Given that, by assumption,the ability to raise children is exogenous and constant over time, we simply denote it via the superscript k rather than introducing a further variable.

being a high human capital agent, is a function of the parents' type jk, the time n_t^{jk} parents dedicate to child care and the quality of child care services e_t, which individuals take as given and which is a choice variable for the government.

The time constraints subject to which agents maximize their objective function are the following:

$$1 = l_t^{jk} + n_t^{jk} + z_t^{jk} \tag{2}$$

$$\overline{a} = n_t^{jk} + d_t^{jk} \tag{3}$$

with l_t^{jk} indicating the labor supply, d_t^{jk} the time spent in day care centers and with $\overline{a} \leq 1$ indicating the care time required by each child. Hereafter we will assume for simplicity that $\overline{a} = 1$.

We assume that for any n_t the following condition holds: $\pi_n^{11}(n_t) < \pi_n^{jk}(n_t) < \pi_n^{22}(n_t)$, with $j \neq k$. These inequalities imply that the time spent with children by an agent who has low ability to raise children and low market ability is less productive in terms of increases in the probability of becoming a skilled individual on the market than the time spent with children by an agent who has high ability both at home and on the market. The productivity of the time devoted to kids by an agent of type 12 and 21 is intermediate when compared to a type 11 and a type 22. We also assume that $\pi_n^{11}(n_t) \leq 0$ and $\pi_n^{22}(n_t) \geq 0$: more time spent with the parent will not increase (resp.: will not decrease) the probability of becoming a high-market ability type if the parent has (high) low ability both at raising kids and on the market. We do not make any assumption on the signs of $\pi_n^{jk}(n_t)$, with $j \neq k$, and consider all the possible cases when discussing the solution of the government maximisation problem.

2.2 Output

Output Y_t is produced according to the following function:

$$Y_t = A \left[\left(f_t^{11} l_t^{11} + f_t^{12} l_t^{12} \right) H^1 + \left(f_t^{21} l_t^{21} + f_t^{22} l_t^{22} \right) H^2 \right] \tag{4}$$

where f^{jk} is the fraction of people of type jk and $A > 0$ is a parameter. Total population is normalized to 1 and the population growth rate is equal to 0.

2.3 Evolution of skills' distribution

The dynamics of the fraction of high market ability people is described by the following linear first order difference equation:

$$f_{t+1}^{2}=\sum_{j=1}^{2}\sum_{k=1}^{2}\pi^{jk}(n_t^{jk})\cdot f_t^{jk} \tag{5}$$

For the fraction of low skilled we have:

$$f_{t+1}^{1}=\sum_{j=1}^{2}\sum_{k=1}^{2}\left[1-\pi^{jk}(n_t^{jk})\right]\cdot f_t^{jk}. \tag{6}$$

Notice that by f_{t+1}^{2} and f_{t+1}^{1} we denote the fraction of high and low market ability individuals, that is $f_{t+1}^{2}\equiv f_{t+1}^{21}+f_{t+1}^{22}$ and $f_{t+1}^{1}\equiv f_{t+1}^{11}+f_{t+1}^{12}$. We assume that the proportion of agents with high or low ability to raise children over the total population is time invariant, that is $f_{t+1}^{21}+f_{t+1}^{11}$ and $f_{t+1}^{22}+f_{t+1}^{12}$ are constant over time. This assumption, along with equations (5) and (6), determine f_t^{jk} for any j,k and t.

2.4 Government

As to the government, the objective function is:

$$W=\sum_{t=0}^{\infty}\rho^t\sum_{j=1}^{2}\sum_{k=1}^{2}f_t^{jk}\cdot\left\{u\left(c_t^{jk},z_t^{jk},n_t^{jk}\right)+\varepsilon\eta\left(\pi^{jk}(n_t^{jk})H^2+(1-\pi^{jk}(n_t^{jk}))H^1\right)\right\} \tag{7}$$

The parameter $\varepsilon\in[0,1]$ allows the preferences of the government to deviate from those of agents, namely, the government may disregard the warm-glow component of individual utility.[4]

As to the budget constraints, to specify them we have to distinguish between the case of a linear tax system and the case of a mixed tax system.

3 Linear tax system

A linear tax system is defined as a system where commodity purchases are taxed according to a set of differentiated proportional taxes and earned income is taxed according to a linear tax (consisting of a uniform marginal income tax rate plus a demogrant). Since labor is the only

[4] How should the warm-glow component factor into social welfare calculations is a philosophical question as much as it is an economic one. According to Hammond (1987), Andreoni (2006) and Diamond (2006), all social welfare prescriptions should be made without counting warm-glow, but should be constrained by behavior that is dictated by seeking warm-glow. For a contrasting view, see Kaplow(1998, 1995).

source of income and a uniform tax on all commodities is equivalent to a proportional tax on labor income, a linear tax system can be equivalently defined as a system where agents receive (pay) a uniform lump-sum subsidy (tax) and commodity purchases are taxed according to a set of differentiated proportional taxes.[5]

We write the agents' budget constraint as:

$$(1+\tau_t^c)c_t^{jk} + (p(e_t)+\tau_t^d)d_t^{jk} = wH^j l_t^{jk} + \Psi_t \tag{8}$$

where the price of consumption is normalized to 1, $p(e_t)$ is the net price of goods d_t, e_t captures the quality of child care services which is taken as given by the individuals, w is the wage in efficiency units, Ψ_t denotes a lump-sum transfer and τ_t^x with $x = c, d$ denotes the tax/subsidy on good x.

The budget constraint for the government can be written as:

$$\tau_t^c \sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk} c_t^{jk} + \tau_t^d \sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk} d_t^{jk} = \Psi_t. \tag{9}$$

The government's budget balances year by year without recurring to debt.

3.1 Solution of the consumer optimization problem and indirect utility function

The maximization of (1) subject to (8), (2) and (3) delivers the following first order conditions for the individual problem:

$$u'_{c_t^{jk}} = \xi_t^{jk}(1+\tau_t^c) \tag{10}$$

$$-u'_{d_t^{jk}} - \eta_t'^{jk}\frac{\partial \pi^{jk}}{\partial n_t^{jk}}(H^2 - H^1) = \xi_t^{jk}(p(e_t)+\tau_t^d - wH^j) \tag{11}$$

$$u'_{z_t^{jk}} = \xi_t^{jk} w H^j, \tag{12}$$

where $u'_{y_t} = \frac{\partial u}{\partial y_t}$ and where ξ_t^{jk} denotes the marginal utility of income for an agent of type jk at time t.

We define V_t^{jk} as the indirect utility function of agent jk at time t, with $V_t^{jk} = u\left(c_t^{jk*}, z_t^{jk*}, n_t^{jk*}\right) + \eta\left(\pi^{jk}(n_t^{jk*})H^2 + (1-\pi^{jk}(n_t^{jk*}))H^1\right)$, where $c_t^{jk*}, z_t^{jk*}, n_t^{jk*}$ are defined by the first order conditions (10), (11), (12) and by the time constraints (2) and (3).

[5]See for instance Atkinson and Stiglitz (1976).

3.2 Solution of the government optimization problem

The government maximizes:

$$\begin{aligned}
\pounds &= \sum_{t=0}^{\infty}\rho^t\sum_{j=1}^{2}\sum_{k=1}^{2}\left[V_t^{jk}-(1-\varepsilon)\,\eta\left(\pi^{jk}\left(n_t^{jk}\right)H^2+\left(1-\pi^{jk}\left(n_t^{jk}\right)\right)H^1\right)\right]f_t^{jk}+ \\
&\quad \sum_{t=0}^{\infty}\rho^t\mu_t\left(\tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2}f_t^{jk}c_t^{jk}+\tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2}f_t^{jk}d_t^{jk}-\Psi_t\right) \\
&\quad -\sum_{t=0}^{\infty}\rho^t\gamma_t\left[f_{t+1}^2-\sum_{j=1}^{2}\sum_{k=1}^{2}f_t^{jk}\pi^{jk}\left(n_t^{jk}\right)\right]
\end{aligned} \tag{13}$$

with respect to τ_t^i and Ψ_t. For the moment, we assume that the quality of child care services e_t is exogenously given and constant at the level $\overline{e}$.

The first order condition with respect to the lump-sum transfer Ψ_t reads as follows:

$$\begin{aligned}
&\sum_{j=1}^{2}\sum_{k=1}^{2}f_t^{jk}\left[\xi_t^{jk}-(1-\varepsilon)\eta_t^{\prime jk}\frac{\partial\pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial\Psi_t}(H^2-H^1)\right]+ \\
&\mu_t\left[\tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2}f_t^{jk}\frac{\partial c_t^{jk}}{\partial\Psi_t}+\tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2}f_t^{jk}\frac{\partial d_t^{jk}}{\partial\Psi_t}-1\right]+ \\
&\gamma_t\sum_{j=1}^{2}\sum_{k=1}^{2}f_t^{jk}\frac{\partial\pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial\Psi_t} \\
=\;& 0.
\end{aligned} \tag{14}$$

where $\eta_t^{\prime jk}$ $(j=1,2;\ k=1,2)$ is defined as $\eta_t^{\prime jk}=\eta'\left(\pi^{jk}\left(n_t^{jk}\right)H^2+\left(1-\pi^{jk}\left(n_t^{jk}\right)\right)H^1\right)$. The above first order condition (14) can be rewritten as:

$$E(b_t^{jk})=1,$$

where $b_t^{jk}=\frac{\xi_t^{jk}}{\mu_t}+\tau_t^c\frac{\partial c_t^{jk}}{\partial\Psi_t}+\tau_t^d\frac{\partial d_t^{jk}}{\partial\Psi_t}-\frac{\partial\pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial\Psi_t}\frac{1}{\mu_t}\left[(1-\varepsilon)(H^2-H^1)\eta_t^{\prime jk}-\gamma_t\right]$ indicates the net marginal social evaluation of agent jk's income. The first term captures the impact that a change in income determined by the lump-sum transfer has on the individual's indirect utility function. The second and third terms indicate the impact on the government revenues associated with the change in the demand functions of the two goods. The fourth term shows the impact that a change in the lump-sum transfer has on the demand for child care and therefore on the probability for agent jk of having a high-market-ability child. If $\varepsilon=1$, the social evaluation of turning a low-market ability into a high-market ability is given by γ_t. When $\varepsilon\neq 1$, the social evaluation will also depend on the degree of laundering out.

We now turn to the first order conditions with respect to τ_t^x $(x = c, d)$. We have:

$$\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left[-\xi_t^{jk}x_t^{jk} - (1-\varepsilon)\eta_t^{\prime jk}\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial \tau_t^x}(H^2 - H^1)\right] +$$
$$\mu_t\left[\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}x_t^{jk} + \tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial c_t^{jk}}{\partial \tau_t^x} + \tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial d_t^{jk}}{\partial \tau_t^x}\right] +$$
$$\gamma_t\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial \tau_t^x} = 0, \qquad (15)$$

for $x = c, d$. Using the Slutsky equation and denoting Hicksian demands by a "tilde", we can rewrite (15), after rearranging terms, as follows:

$$\frac{\tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\partial\widetilde{x}_t^{jk}/\partial\tau_t^x + \tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\partial\widetilde{x}_t^{jk}/\partial\tau_t^x}{x_t}$$
$$= \quad -\left[1 - Cov(b_t^{jk}, \frac{x_t^{jk}}{x_t})\right] +$$
$$\frac{1}{\mu_t}\sum_{j=1}^{2}\sum_{k=1}^{2}\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial\widetilde{d}_t^{jk}/\partial\tau_t^x}{x_t} f_t^{jk}\left[(1-\varepsilon)(H^2 - H^1)\eta_t^{\prime jk} - \gamma_t\right], \qquad (16)$$

for $x = c, d$. The proportional change in the aggregate compensated demand for good x due to indirect taxes is determined by two terms. The first term on the right hand side is entirely standard and it captures the government redistributive concerns. The higher is $Cov(b_t^{jk}, \frac{x_t^{jk}}{x_t})$, the lower should be the reduction of the consumption of good x due to the tax system. The second one is the new term stemming from the impact that day care arrangements have on human capital accumulation. In this new term we can identify two components: the first one, that is $\frac{1}{\mu_t}\sum_{j=1}^{2}\sum_{k=1}^{2}\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial\widetilde{d}_t^{jk}/\partial\tau_t^x}{x_t} f_t^{jk}\left[(1-\varepsilon)(H^2 - H^1)\eta_t^{\prime jk}\right]$, depends on whether the government takes into account fully $(\varepsilon = 1)$, partially $(0 < \varepsilon < 1)$ or not at all $(\varepsilon = 0)$ the warm-glow component of individual preferences. The second one, namely $-\frac{1}{\mu_t}\sum_{j=1}^{2}\sum_{k=1}^{2}\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial\widetilde{d}_t^{jk}/\partial\tau_t^x}{x_t} f_t^{jk}\gamma_t$, identifies the externality related to the assumption of imperfect altruism. Notice that these two terms always push in the opposite direction provided that $\varepsilon \neq 1$. As to the sign of the overall correction, notice that the instruments available to the government are linear, which implies that the tax rates applied to the two goods need to be the same irrespective of the market ability and of the ability to raise children. For this reason, tax rates have to average the adjustments ideally required to correct the behavior of the four types of agents. Here the sign of $\partial\pi^{jk}/\partial d_t^{jk}$ becomes relevant. We have assumed, as it is natural, that $\partial\pi^{11}/\partial d_t^{11} > 0$ while $\partial\pi^{22}/\partial d_t^{22} < 0$: the correction ideally imposed on these two types is therefore of opposite sign, at least when the expression within square brackets in the last line of (16) takes the same sign for agents of type 11 and agents of type 22. Thus, if we consider for example the case of $\varepsilon = 1$, we will have that the adjustment ideally required to correct the time-allocation choice of a parent of type 11 (resp.: 22) would call for encouraging (resp.: discouraging) the consumption of day care services and of their Hicksian complements.

If intuition leads quite naturally to assume that $\partial\pi^{11}/\partial d_t^{11} > 0$ and $\partial\pi^{22}/\partial d_t^{22} < 0$, it is harder to make assumptions on the sign of the other two derivatives, namely $\partial\pi^{jk}/\partial d_t^{jk}$ when $j \neq k$. It seems reasonable to believe that the sign of the derivative will in these cases crucially depend on the value of d (and therefore on the number of hours spent by the parent with the child) at which it is evaluated. Specifically, it seems reasonable to assume that the sign of the derivative is positive for low values of d (high values of n) and it becomes negative for values of d exceeding (values of n below) a given threshold, which will in general be different for parents of type 12 and parents of type 21.[6] For this reason, we have decided to avoid making here any specific assumption on the sign of $\partial\pi^{jk}/\partial d_t^{jk}$ for $j \neq k$. Therefore, depending on the interaction between the sign taken by each of these derivatives and the sign of the corresponding expression trading off the laundering- and the externality-component (namely $(1-\varepsilon)(H^2 - H^1)\eta_t'^{jk} - \gamma_t$ for $j \neq k$), the formula characterizing the tax rate on good x will incorporate terms referring to agents of type jk, with $j \neq k$, pushing either in the direction of encouraging or in the direction of discouraging the consumption of good x.

We now consider the case where the government can also set the quality of child care e_t.

[6] We do not deny that a similar pattern might also hold for agents of type 11 and 22. What we have in mind is that for agents of type 11 the derivative of π^{11} with respect to d^{11} becomes negative for values of d^{11} which are so large that they cannot prevail at a social optimum; similarly, for agents of type 22 we admit the theoretical possibility that the derivative of π^{22} with respect to d^{22} is positive but we implicitly assume that this can only happen for values of d^{22} which are so small that they cannot be compatible with their labor supply at a social optimum.

Differentiating (13) with respect to e_t, we find:

$$\begin{aligned}
&\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left[\frac{\partial V_t^{jk}}{\partial e_t} - (1-\varepsilon)\,\eta_t^{\prime jk}\left(\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial e_t} + \frac{\partial \pi^{jk}}{\partial e_t}\right)\left(H^2 - H^1\right)\right] + \qquad (17)\\
&\mu_t\left(\tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial c_t^{jk}}{\partial e_t} + \tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial d_t^{jk}}{\partial e_t}\right) + \\
&\gamma_t\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left(\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial e_t} + \frac{\partial \pi^{jk}}{\partial e_t}\right) + p'(e_t)\,\Upsilon_t \\
= \;& 0,
\end{aligned}$$

where Υ_t has been defined as:

$$\begin{aligned}
\Upsilon_t \equiv\; &\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left[\frac{\partial V_t^{jk}}{\partial q_t} - (1-\varepsilon)\left(H^2 - H^1\right)\eta_t^{\prime jk}\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial q_t}\right] + \qquad (18)\\
&\mu_t\left(\tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial c_t^{jk}}{\partial q_t} + \tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial d_t^{jk}}{\partial q_t}\right) + \\
&\gamma_t\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial q_t}.
\end{aligned}$$

with $(p(e_t) + \tau_t^d) = q_t$. Using the first order condition with respect to τ_t^d (15), it is straightforward to conclude that $\Upsilon_t = -\mu_t d_t$, where $d_t = \sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk} d_t^{jk}$ denotes aggregate consumption of day-care. We can therefore rewrite (17) as:

$$\begin{aligned}
\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial V_t^{jk}}{\partial e_t} = \;& p'(e_t)\,\mu_t d_t - \mu_t\left(\tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial c_t^{jk}}{\partial e_t} + \tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial d_t^{jk}}{\partial e_t}\right) \\
&+\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left[(1-\varepsilon)\,\eta_t^{\prime jk}\left(H^2 - H^1\right) - \gamma_t\right]\left(\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial e_t} + \frac{\partial \pi^{jk}}{\partial e_t}\right) \qquad (19)
\end{aligned}$$

The left-hand side indicates the sum of the changes in indirect utilities due to an increase in the quality of day care. The first term on the right hand side captures the changes in the cost born by the agents as a consequence of the increase in quality, for a given demand of day care services. The second term measures the impact on government revenues of a higher quality of day care. The last term takes into account that the change in quality influences the probability of becoming skilled both directly (the term $\partial \pi^{jk}/\partial e_t$) and indirectly (the term $\left(\partial \pi^{jk}/\partial d_t^{jk}\right)\left(\partial d_t^{jk}/\partial e_t\right)$). The implied correction depends, as above, on the presence or absence of laundering out in the social welfare function and on the intergenerational externality stemming from imperfect altruism.

4 Mixed tax system

Given that the choice of the optimal commodity tax structure boils down in our two-good model to the choice of the optimal tax rate on expenses for day care services, we can safely skip superscripts and denote by τ_t the commodity tax (or subsidy) that applies at time t on expenses for day care services.

Since the government can observe earned income at an individual level but it can observe neither an individual's labor supply nor his wage rate, the design of the nonlinear income tax is constrained by a set of self-selection constraints. These constraints require that each agent must prefer the point on the income tax schedule intended for his type rather than misrepresent his true ability type and choose a point intended for some other types. An agent misrepresenting his ability type is called a mimicker. To avoid dealing with well-known problems related to multidimensional screening, we will assume in this section that there is perfect correlation between the two types of abilities, market ability and ability to raise children. Under the assumption of perfect correlation, the four type model that we have analyzed in the previous section boils down to a two type model where the population is divided between agents of type 11 and agents of type 22. Since there are no longer agents of type jk, with $j \neq k$, we can safely simplify the notation and use a single index to distinguish between different agents. Thus, we will hereafter use the index 1 to denote low-ability agents and the index 2 to denote high-ability agents. In our analysis we will focus on the so-called normal case where the only binding self-selection constraint is the one ruling out the possibility that high-skilled agents mimic low-skilled ones. Defining B_t^j as $B_t^j \equiv Y_t^j - T_t\left(Y_t^j\right)$, the government's problem can be equivalently stated as the problem of offering at each time t two different bundles in the (Y, B)-space, one for the high-skilled and one for the low-skilled, subject to a self-selection and a public budget constraint.

The design problem can be therefore summarized by the Lagrangian:

$$\begin{aligned} \pounds \;=\; & \sum_{t=0}^{\infty}\rho^t\sum_{j=1}^{2}\left[V_t^j-(1-\varepsilon)\,\eta\left(\pi^j\left(n_t^j\right)H^2+\left(1-\pi^j\left(n_t^j\right)\right)H^1\right)\right]f_t^j+ \\ & \sum_{t=0}^{\infty}\rho^t\mu_t\sum_{j=1}^{2}\left(Y_t^j-B_t^j+\tau_t d_t^j\right)f_t^j-\sum_{t=0}^{\infty}\rho^t\gamma_t\left[f_{t+1}^2-\sum_{j=1}^{2}f_t^j\pi^j\left(n_t^j\right)\right]+ \\ & \sum_{t=0}^{\infty}\rho^t\lambda_t\left(V_t^2-\widehat{V}_t^2\right), \end{aligned} \qquad (20)$$

where a "hat" is used to indicate a variable that pertains to a mimicker.

The value of e_t is for the moment taken as exogenously given and time-invariant. We will relax this assumption later on.

The first order conditions for Y_t^2 and B_t^2 are:

$$\left(f_t^2+\lambda_t\right)\frac{\partial V_t^2}{\partial Y_t^2}=\left\{\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^2-\gamma_t\right]\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial Y_t^2}-\mu_t\left(1+\tau_t\frac{\partial d_t^2}{\partial Y_t^2}\right)\right\}f_t^2; \qquad (21)$$

$$\left(f_t^2+\lambda_t\right)\frac{\partial V_t^2}{\partial B_t^2}=\left\{\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^2-\gamma_t\right]\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right\}f_t^2. \qquad (22)$$

Dividing (21) by (22) and multiplying the result by the right hand side of (22), we get:

$$\begin{aligned}&\frac{\frac{\partial V_t^2}{\partial Y_t^2}}{\frac{\partial V_t^2}{\partial B_t^2}}\left\{\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^2-\gamma_t\right]\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right\}\\ =\;&\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^2-\gamma_t\right]\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial Y_t^2}-\mu_t\left(1+\tau_t\frac{\partial d_t^2}{\partial Y_t^2}\right). \qquad (23)\end{aligned}$$

Since from the optimization problem solved by the high-skilled agents we can implicitly express the marginal tax rate faced by them as $T'\left(Y_t^2\right)=1+\left(\frac{\partial V_t^2}{\partial Y_t^2}/\frac{\partial V_t^2}{\partial B_t^2}\right)$, collecting terms in (23) gives:

$$T'\left(Y_t^2\right)=-\left(\frac{dd_t^2}{dY_t^2}\right)_{dV_t^2=0}\tau_t+\frac{1}{\mu_t}\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^2-\gamma_t\right]\frac{\partial \pi^2}{\partial n_t^2}\left(\frac{dn_t^2}{dY_t^2}\right)_{dV_t^2=0}, \qquad (24)$$

where $\left(\frac{dd_t^2}{dY_t^2}\right)_{dV_t^2=0}\equiv\frac{\partial d_t^2}{\partial Y_t^2}+MRS_t^2\frac{\partial d_t^2}{\partial B_t^2}\equiv\frac{\partial d_t^2}{\partial Y_t^2}-\left(\frac{\partial V_t^2}{\partial Y_t^2}/\frac{\partial V_t^2}{\partial B_t^2}\right)\frac{\partial d_t^2}{\partial B_t^2}$. Finally, notice that $d_t^2=1-n_t^2=l_t^2+z_t^2=\frac{Y_t^2}{wH^2}+z_t^2$. Therefore $\left(\frac{dd_t^2}{dY_t^2}\right)_{dV_t^2=0}=\frac{1}{wH^2}+\left(\frac{dz_t^2}{dY_t^2}\right)_{dV_t^2=0}$ and $\left(\frac{dd_t^2}{dY_t^2}\right)_{dV_t^2=0}=-\left(\frac{dn_t^2}{dY_t^2}\right)_{dV_t^2=0}$. From (24) we can easily calculate the marginal effective tax rate faced by high skilled agents. Let's denote it by $METR_t^2$. This is defined as $T'\left(Y_t^2\right)+\left(\frac{dd_t^2}{dY_t^2}\right)_{dV_t^2=0}\tau_t$ and therefore we have:

$$METR_t^2=\frac{1}{\mu_t}\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^2-\gamma_t\right]\frac{\partial \pi^2}{\partial n_t^2}\left(\frac{dn_t^2}{dY_t^2}\right)_{dV_t^2=0}. \qquad (25)$$

We know that $\partial\pi^2/\partial n_t^2>0$, namely that additional time spent by high-skilled agents with their children increases the probability that, as adults, they will be high-skilled too. Under the reasonable assumption that $\left(dn_t^2/dY_t^2\right)_{dV_t^2=0}<0$ (since additional time devoted to working implies that the total amount of time that can be allocated on z and n goes down), the sign of (25) is the opposite of the sign of the term within square brackets.

When the government respects the individuals' preferences, so that $\varepsilon = 1$, the METR faced by the high skilled agents is therefore positive, implying that the overall effect of the tax system is to induce high-skilled agents to under-provide labor supply in order to spend more time with their children. This is required in order to induce the high-skilled adults at time t to internalize the social welfare effect generated by the link between their time allocation decision and the proportion of high-skilled adults at time $t+1$. Spending more time with their children, the high-skilled agents raise the probability that, growing up, their children will become high-skilled adults.

If however the government launders, fully ($\varepsilon = 0$) or partially ($0 < \varepsilon < 1$), the individuals' preferences into the social welfare function, one cannot rule out the possibility that the METR faced by the high-skilled agents turns out to be negative. The reason is that, as ε becomes smaller, the need to provide high-skilled agents with incentives to spend more time with their children is weakened due to the fact that, from the government's point of view, high-skilled agents overvalue the utility that they get from spending time with their children. As ε approaches zero, this effect might become so strong that, even if additional time spent by high-skilled parents with their children raises the probability of these becoming high-skilled as adults, from a social point of view parents appear to be over-investing in time spent with their children. To correct for this, a negative marginal effective tax rate on high-skilled agents might be warranted as an indirect instrument to induce agents to work more and reduce total time spent with their children.

Consider now the first order conditions for Y_t^1 and B_t^1. These are respectively given by:

$$f_t^1 \frac{\partial V_t^1}{\partial Y_t^1} = \lambda_t \frac{\partial \widehat{V}_t^2}{\partial Y_t^1} + \left\{ \left[(1-\varepsilon) \left(H^2 - H^1 \right) \eta_t'^1 - \gamma_t \right] \frac{\partial \pi^1}{\partial n_t^1} \frac{\partial n_t^1}{\partial Y_t^1} - \mu_t \left(1 + \tau_t \frac{\partial d_t^1}{\partial Y_t^1} \right) \right\} f_t^1; \tag{26}$$

$$f_t^1 \frac{\partial V_t^1}{\partial B_t^1} = \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1} + \left\{ \left[(1-\varepsilon) \left(H^2 - H^1 \right) \eta_t'^1 - \gamma_t \right] \frac{\partial \pi^1}{\partial n_t^1} \frac{\partial n_t^1}{\partial B_t^1} - \mu_t \left(-1 + \tau_t \frac{\partial d_t^1}{\partial B_t^1} \right) \right\} f_t^1. \tag{27}$$

Dividing (26) by (27) and multiplying the result by the right hand side of (27), we get:

$$\begin{aligned} & \frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}} \left\{ \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1} + \left[\left((1-\varepsilon) \left(H^2 - H^1 \right) \eta_t'^1 - \gamma_t \right) \frac{\partial \pi^1}{\partial n_t^1} \frac{\partial n_t^1}{\partial B_t^1} - \mu_t \left(-1 + \tau_t \frac{\partial d_t^1}{\partial B_t^1} \right) \right] f_t^1 \right\} \\ = \; & \lambda_t \frac{\partial \widehat{V}_t^2}{\partial Y_t^1} + \left\{ \left[(1-\varepsilon) \left(H^2 - H^1 \right) \eta_t'^1 - \gamma_t \right] \frac{\partial \pi^1}{\partial n_t^1} \frac{\partial n_t^1}{\partial Y_t^1} - \mu_t \left(1 + \tau_t \frac{\partial d_t^1}{\partial Y_t^1} \right) \right\} f_t^1. \end{aligned} \tag{28}$$

Since from the optimization problem solved by the low-skilled agents we can implicitly express the marginal tax rate faced by them as $T'\left(Y_t^1\right) = 1 + \left(\frac{\partial V_t^1}{\partial Y_t^1} / \frac{\partial V_t^1}{\partial B_t^1} \right)$, collecting terms in

(28) gives:

$$T'\left(Y_t^1\right) = -\left(\frac{dd_t^1}{dY_t^1}\right)_{dV_t^1=0}\tau_t + \frac{\lambda_t}{\mu_t f_t^1}\frac{\partial\widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial\widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial\widehat{V}_t^2}{\partial B_t^1}} - \frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\right) + \frac{1}{\mu_t}\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^1 - \gamma_t\right]\frac{\partial\pi^1}{\partial n_t^1}\left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0}, \tag{29}$$

where $\left(\frac{dd_t^1}{dY_t^1}\right)_{dV_t^1=0} \equiv \frac{\partial d_t^1}{\partial Y_t^1} + MRS_t^1\frac{\partial d_t^1}{\partial B_t^1} \equiv \frac{\partial d_t^1}{\partial Y_t^1} - \left(\frac{\partial V_t^1}{\partial Y_t^1}/\frac{\partial V_t^1}{\partial B_t^1}\right)\frac{\partial d_t^1}{\partial B_t^1}$. From (29) we can easily calculate the marginal effective tax rate faced by low skilled agents. Let's denote it by $METR_t^1$. This is defined as $T'\left(Y_t^1\right) + \left(\frac{dd_t^1}{dY_t^1}\right)_{dV_t^1=0}\tau_t$ and therefore we have:

$$METR_t^1 = \frac{\lambda_t}{\mu_t f_t^1}\frac{\partial\widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial\widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial\widehat{V}_t^2}{\partial B_t^1}} - \frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\right) + \frac{1}{\mu_t}\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^1 - \gamma_t\right]\frac{\partial\pi^1}{\partial n_t^1}\left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0}. \tag{30}$$

The second term on the right hand side of (30) has the same structure of the term appearing on the right hand side of (25) and can be interpreted in a similar way. The only thing that differs is that $\partial\pi^1/\partial n_t^1 < 0$, whereas in (25) we had $\partial\pi^2/\partial n_t^2 > 0$. This reflects our assumption that the probability that children of low-skilled agents become high-skilled adults is negatively affected when they spend more time with their parents and less time in day care centers.[7] Thus, the sign of the second term on the right hand side of (30) is the same as the sign of the expression within square brackets. In particular, when the government respects the agents' preferences and chooses $\varepsilon = 1$, the second term on the right hand side of (30) tends to reduce the marginal effective tax rate faced by low-skilled agents. This represents a way to induce low-skilled agents to work more and substitute consumption for leisure time (which also includes time spent with the offspring). Being unable to directly control the amount of time that parents devote to their children, the government affects the agents' incentives to engage in labor market activities in order to influence the time they spend with their children and let them internalize the social welfare effect generated by the link between their time allocation decision and the proportion of high-skilled adults at time $t+1$. If however the government launders the agents' preferences in the social welfare function and chooses $0 \leq \varepsilon < 1$, the sign of the second term on the right hand side of (30) might change from negative to positive,

[7]Remember that we have assumed that children must be taken care of all the time, either by parents themselves or at day care centers. Therefore, if time spent with parents goes up, time spent in day care centers necessarily goes down.

reflecting the fact that, from the perspective of the government, low-skilled agents undervalue the utility that they get from spending time with their children.

A general point which is worth noticing is that, although we have assumed that the sign of $\partial\pi^1/\partial n_t^1$ is opposite to the sign of $\partial\pi^2/\partial n_t^2$, the sign of the corrective terms appearing in (25) and (30), namely those depending respectively on $\partial\pi^2/\partial n_t^2$ and $\partial\pi^1/\partial n_t^1$, are not necessarily opposite, unless the government fully respects the agents' preferences ($\varepsilon = 0$). This happens because with $0 < \varepsilon < 1$ it might well be the case that the sign of the term within square brackets in (25) is opposite to that of the term within square brackets in (30). Notice however that we cannot claim that the likelihood of getting the same sign for the corrective terms in (25) and (30) increases launders further the agents' preferences in its objective function, i.e. as ε becomes smaller. The obvious reason is that a change in the value of ε chosen by the government triggers a change in the optimal equilibrium levels of the expected human capital of offspring for the high- and low-skilled parents.

The first term on the right hand side of (30) reflects the distortion that the tax system should impose on the labor supply of the low-skilled agents in order to prevent the high-skilled agents from being tempted to become mimicker and choose the (Y, B)-bundle intended for the low-skilled. It is due to the assumption that the government cannot observe "who is who" and is therefore constrained to design the income tax schedule subject to a (set of) self-selection constraint(s). The sign of this self-selection term coincides with the sign of the expression within brackets. In standard models of nonlinear redistributive income taxation,[8] it is common practice to make the so-called agent-monotonicity assumption. This assumption requires that, at any given point in the (Y, B)-space, the higher the wage rate of an agent, the flatter the indifference curves are. If this assumption is satisfied, we can safely conclude that the self-selection term takes a positive sign and therefore requires a downward distortion on the labor supply of low-skilled agents. Notice however that whereas in standard models of optimal nonlinear taxation it is usually sufficient to assume normality of consumption to get the agent-monotonicity assumption satisfied, in our model this is no longer enough. Intuitively, this is due to the fact that a high-skilled mimicker and a true low-skilled agent do not only differ with respect to their labor supply but in general also with respect to the amount available for private consumption (once expenses on day-care services have been subtracted). To explore this issue in more details, take any given bundle in the (Y, B)-space and consider the marginal rate of substitution

[8] See, for example, Stiglitz (1982) and Edwards, Keen, and Tuomala (1994).

between Y and B for a generic agent of type j.[9] This is given by $-\left(\partial V^j/\partial Y\right)/\left(\partial V^j/\partial B\right)$. Assuming a utility function of the form $u\left(c^j, z^j, n^j\right) + \eta\left(\pi^j\left(n^j\right)H^2 + \left(1-\pi^j\left(n^j\right)\right)H^1\right)$, the conditional indirect utility for an agent of type j, $V^j(Y,B)$, is obtained maximizing $u\left(c^j, z^j, n^j\right) + \eta\left(\pi^j\left(n^j\right)H^2 + \left(1-\pi^j\left(n^j\right)\right)H^1\right)$ subject to the budget constraint $c^j = B - (p(e)+\tau)d^j = B - (p(e)+\tau)\left(l^j + z^j\right) = B - (p(e)+\tau)\left[\left(Y/wH^j\right) + z^j\right]$ and the time constraint $n^j = 1 - l^j - z^j$. This implies that $\partial V^j/\partial B = \partial u\left(\cdot; wH^j\right)/\partial c^j$ and $\partial V^j/\partial Y = -\left(wH^j\right)^{-1}\left[\frac{\partial u\left(\cdot;wH^j\right)}{\partial c^j}(p(e)+\tau) + \frac{\partial u\left(\cdot;wH^j\right)}{\partial n^j} + \left(H^2-H^1\right)\eta'^j\frac{\partial \pi^j}{\partial n^j}\right]$. Thus, we have:

$$
\begin{aligned}
-\frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}} &= \left(wH^1\frac{\partial u\left(\cdot;wH^1\right)}{\partial c_t^1}\right)^{-1}\left[\frac{\partial u\left(\cdot;wH^1\right)}{\partial c_t^1}\left(p(e)+\tau_t\right) + \frac{\partial u\left(\cdot;wH^1\right)}{\partial n_t^1} + \left(H^2-H^1\right)\eta_t'^1\frac{\partial \pi^1}{\partial n_t^1}\right] \\
&= \frac{p(e)+\tau_t}{wH^1} + \frac{\frac{\partial u\left(\cdot;wH^1\right)}{\partial n_t^1} + \left(H^2-H^1\right)\eta_t'^1\frac{\partial \pi^1}{\partial n_t^1}}{wH^1\frac{\partial u\left(\cdot;wH^1\right)}{\partial c_t^1}} \qquad (31)
\end{aligned}
$$

$$
\begin{aligned}
-\frac{\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} &= \left(wH^2\frac{\partial \widehat{u}\left(\cdot;wH^2\right)}{\partial \widehat{c}_t^2}\right)^{-1}\left[\frac{\partial \widehat{u}\left(\cdot;wH^2\right)}{\partial \widehat{c}_t^2}\left(p(e)+\tau_t\right) + \frac{\partial \widehat{u}\left(\cdot;wH^2\right)}{\partial \widehat{n}_t^2} + \left(H^2-H^1\right)\widehat{\eta}_t'^2\frac{\partial \widehat{\pi}^2}{\partial \widehat{n}_t^2}\right] \\
&= \frac{p(e)+\tau_t}{wH^2} + \frac{\frac{\partial \widehat{u}\left(\cdot;wH^2\right)}{\partial \widehat{n}_t^2} + \left(H^2-H^1\right)\widehat{\eta}_t'^2\frac{\partial \widehat{\pi}^2}{\partial \widehat{n}_t^2}}{wH^2\frac{\partial \widehat{u}\left(\cdot;wH^2\right)}{\partial \widehat{c}_t^2}}. \qquad (32)
\end{aligned}
$$

Comparing (31) with (32), it is obvious that the first term on the right hand side of the latter is smaller than the corresponding term on the right side of the former. This contributes to make the marginal rate of substitution for the mimicker smaller than the marginal rate of substitution for a true low-skilled. It is also true, however, that the labor supply provided by a mimicker is smaller than the labor supply provided by a true low-skilled, which means that a mimicker has a larger total amount of time to devote to leisure activities. Thus, if a mimicker spends more time with his kid than a true low-skilled,[10] and given that $d_t^j = 1 - n_t^j$, the expenses for day-care services will be smaller for a mimicker than for a true low-skilled. This in turn means that $\widehat{c}_t^2 > c_t^1$ and therefore that $\partial \widehat{u}\left(\cdot; wH^2\right)/\partial \widehat{c}_t^2$ (appearing at the denominator of the second term on the right side of (32)) might be smaller than $\partial u\left(\cdot; wH^1\right)/\partial c_t^1$ (appearing at the denominator of the second term on the right side of (31)). Moreover, it is also true

[9] We will here for simplicity suppress the time subscripts.

[10] The available evidence seems to support the idea that there is a positive wage elasticity for time spent with children and a negative wage elasticity for time spent on other leisure activities. See e.g. Kimmel and Connelly (2007) and Guryan, Hurst, and Kearney (2008).

that $\left(H^2-H^1\right)\widehat{\eta}_t'^2\partial\widehat{\pi}^2/\partial\widehat{n}_t^2>0$ whereas $\left(H^2-H^1\right)\eta_t'^1\partial\pi^1/\partial n_t^1<0$ and this also represents an effect contributing to raise the marginal rate of substitution of the mimicker relatively to that of a true low-skilled. This, on the other hand, has to be weighed against the fact that, as a mimicker spends more time with his kid than a true low-skilled, $\partial\widehat{u}\left(\cdot;wH^2\right)/\partial\widehat{n}_t^2$ tends to be smaller than $\partial u\left(\cdot;wH^1\right)/\partial n_t^1$.

Let's investigate now how the value for τ_t should be optimally selected. The first order condition for τ_t is given by:

$$\begin{aligned}&\sum_{j=1}^{2}\left[\frac{\partial V_t^j}{\partial\tau_t}-(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^j\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial\tau_t}\right]f_t^j+\mu_t\sum_{j=1}^{2}\left(d_t^j+\tau_t\frac{\partial d_t^j}{\partial\tau_t}\right)f_t^j+\lambda_t\left(\frac{\partial V_t^2}{\partial\tau_t}-\frac{\partial\widehat{V}_t^2}{\partial\tau_t}\right)\\&=-\gamma_t\sum_{j=1}^{2}f_t^j\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial\tau_t}.\end{aligned}\tag{33}$$

Using the identity $\frac{\partial V_t^j}{\partial\tau_t}=-d_t^j\frac{\partial V_t^j}{\partial B_t^j}$ we can rewrite the equation above as:

$$\begin{aligned}&\sum_{j=1}^{2}\left[-d_t^j\frac{\partial V_t^j}{\partial B_t^j}-(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^j\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial\tau_t}\right]f_t^j+\mu_t\sum_{j=1}^{2}\left(d_t^j+\tau_t\frac{\partial d_t^j}{\partial\tau_t}\right)f_t^j+\\&\lambda_t\left(-d_t^2\frac{\partial V_t^2}{\partial B_t^2}+\widehat{d}_t^2\frac{\partial\widehat{V}_t^2}{\partial B_t^1}\right)\\&=-\gamma_t\sum_{j=1}^{2}f_t^j\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial\tau_t}.\end{aligned}\tag{34}$$

Multiplying (22) and (27) by respectively d_t^2 and d_t^1, we can find the following two expressions for $-\left(f_t^2+\lambda_t\right)d_t^2\frac{\partial V_t^2}{\partial B_t^2}$ and $-d_t^1\frac{\partial V_t^1}{\partial B_t^1}$:

$$-\left(f_t^2+\lambda_t\right)\frac{\partial V_t^2}{\partial B_t^2}d_t^2=-\left\{\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^2-\gamma_t\right]\frac{\partial\pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right\}f_t^2d_t^2;\tag{35}$$

$$-f_t^1d_t^1\frac{\partial V_t^1}{\partial B_t^1}=-\lambda_t d_t^1\frac{\partial\widehat{V}_t^2}{\partial B_t^1}-\left\{\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^1-\gamma_t\right]\frac{\partial\pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1}-\mu_t\left(-1+\tau_t\frac{\partial d_t^1}{\partial B_t^1}\right)\right\}f_t^1d_t^1.\tag{36}$$

Substituting (35) and (36) into (34) and using the Slutsky-type decomposition $\frac{\partial d_t^j}{\partial\tau_t}=\frac{\partial\widetilde{d}_t^j}{\partial\tau_t}-d_t^j\frac{\partial d_t^j}{\partial B_t^j}$ gives:

$$
\begin{aligned}
& -\left\{\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^2-\gamma_t\right]\frac{\partial\pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right\}f_t^2 d_t^2 + \\
& -\lambda_t d_t^1\frac{\partial \widehat{V}_t^2}{\partial B_t^1}-\left\{\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^1-\gamma_t\right]\frac{\partial\pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1}-\mu_t\left(-1+\tau_t\frac{\partial d_t^1}{\partial B_t^1}\right)\right\}f_t^1 d_t^1 + \\
& \mu_t\sum_{j=1}^{2}\left[d_t^j+\tau_t\left(\frac{\partial\widetilde{d}_t^j}{\partial\tau_t}-d_t^j\frac{\partial d_t^j}{\partial B_t^j}\right)\right]f_t^j-\sum_{j=1}^{2}(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^j\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial\tau_t}f_t^j+\lambda_t\widehat{d}_t^2\frac{\partial\widehat{V}_t^2}{\partial B_t^1} \\
= \; & -\gamma_t\sum_{j=1}^{2}f_t^j\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial\tau_t}. \qquad (37)
\end{aligned}
$$

Simplifying terms in (37) gives:

$$
\begin{aligned}
& \sum_{j=1}^{2}\tau_t\frac{\partial\widetilde{d}_t^j}{\partial\tau_t}f_t^j=-\frac{\gamma_t}{\mu_t}\left[\sum_{j=1}^{2}f_t^j\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial\tau_t}+\frac{\partial\pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2}f_t^2d_t^2+\frac{\partial\pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1}f_t^1d_t^1\right]+ \\
& \frac{1}{\mu_t}\sum_{j=1}^{2}(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^j\frac{\partial\pi^j}{\partial n_t^j}\left(\frac{\partial n_t^j}{\partial\tau_t}+\frac{\partial n_t^j}{\partial B_t^j}d_t^j\right)f_t^j+\frac{\lambda_t}{\mu_t}\left[d_t^1-\widehat{d}_t^2\right]\frac{\partial\widehat{V}_t^2}{\partial B_t^1}. \qquad (38)
\end{aligned}
$$

Defining the compensated effect on n_t^j of a marginal increase in τ_t, $\frac{\partial\widetilde{n}_t^j}{\partial\tau_t}$, as $\frac{\partial\widetilde{n}_t^j}{\partial\tau_t}\equiv\frac{\partial n_t^j}{\partial\tau_t}+d_t^j\frac{\partial n_t^j}{\partial B_t^j}$, we can rewrite (38) in a more compact form as:

$$
\begin{aligned}
& \sum_{j=1}^{2}\tau_t\frac{\partial\widetilde{d}_t^j}{\partial\tau_t}f_t^j=\frac{1}{\mu_t}\sum_{j=1}^{2}(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^j\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial\widetilde{n}_t^j}{\partial\tau_t}f_t^j+ \\
& \frac{\lambda_t}{\mu_t}\left[d_t^1-\widehat{d}_t^2\right]\frac{\partial\widehat{V}_t^2}{\partial B_t^1}-\frac{\gamma_t}{\mu_t}\left[\sum_{j=1}^{2}f_t^j\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial\widetilde{n}_t^j}{\partial\tau_t}\right].
\end{aligned}
$$

Finally, exploiting the time-constraint identity $d_t^j=1-n_t^j$, and therefore $\frac{\partial\widetilde{n}_t^j}{\partial\tau_t}=-\frac{\partial\widetilde{d}_t^j}{\partial\tau_t}$ and $\frac{\partial\pi^j}{\partial n_t^j}=-\frac{\partial\pi^j}{\partial d_t^j}$, we end up with:

$$
\tau_t=\frac{\lambda_t\left[d_t^1-\widehat{d}_t^2\right]\frac{\partial\widehat{V}_t^2}{\partial B_t^1}+\sum_{j=1}^{2}\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^j-\gamma_t\right]\frac{\partial\pi^j}{\partial d_t^j}\frac{\partial\widetilde{d}_t^j}{\partial\tau_t}f_t^j}{\mu_t\sum_{j=1}^{2}\frac{\partial\widetilde{d}_t^j}{\partial\tau_t}f_t^j}. \qquad (39)
$$

The denominator of the expression on the right hand side of (39) is negative and it provides a measure of the deadweight loss generated by distortionary commodity taxation. Thus, the sign of τ_t is the opposite of the sign of the numerator of the expression on the right side of (39). The first term at the numerator depends on the difference between the amount of day-care services used by a true low-skilled and a high-skilled mimicker. As we have already previously noticed, a mimicker provides a smaller labor supply than a true low-skilled and it is therefore reasonable to assume that $d_t^1 - \widehat{d}_t^2 > 0$. Thus, the first term on the right side of (39) calls for a subsidy on the purchase of day-care services. Intuitively, the underlying idea is that, given that $d_t^1 > \widehat{d}_t^2$ and starting from a situation where $\tau_t = 0$, it is possible to relax the binding self-selection constraint by introducing a small subsidy to child care expenditures while at the same time leaving the utility of all non-mimicking agents unaffected by raising their income tax payments (lowering B_t^1 and B_t^2) by respectively d_t^1 and d_t^2. To make the interpretation of the second term appearing at the numerator of (39) easier, it is convenient to introduce the variable ζ_t^j, defined as $\zeta_t^j \equiv \frac{\partial \widetilde{d}_t^j}{\partial \tau_t} f_t^j / \sum_{i=1}^{2} \frac{\partial \widetilde{d}_t^i}{\partial \tau_t} f_t^i$, where ζ_t^j represents the normalized change, generated by a marginal increase in τ_t, in the compensated demand by agents of skill type j for day-care services. Thus, we can rewrite τ_t as

$$\tau_t = \frac{\lambda_t}{\mu_t} \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \frac{d_t^1 - \widehat{d}_t^2}{\sum_{j=1}^{2} \frac{\partial \widetilde{d}_t^j}{\partial \tau_t} f_t^j} - \frac{1}{\mu_t} \sum_{j=1}^{2} \left[(1-\varepsilon) \left(H^2 - H^1 \right) \eta_t^{\prime j} - \gamma_t \right] \frac{\partial \pi^j}{\partial n_t^j} \zeta_t^j. \tag{40}$$

Written in this form, the second term on the right hand side of (40) is reminiscent of a similar term appearing in (25) and (30). The main difference is that in (40) we take a sum over $j = 1, 2$ whereas in both (25) and (30) we only have a type-specific term. The reason for this is related to the different degree of sophistication of the available tax instruments. Labor income is assumed to be taxable on the basis of a nonlinear schedule. This implies that, subject to a self-selection constraint, the government can offer agents type-specific marginal income tax rates. Purchases of day-care services, on the other hand, are assumed to be taxable only linearly, meaning that the commodity tax (or subsidy) rate on day-care purchases is the same for all agents, irrespective of the skill type.[11] But for the purpose of letting agents internalize the social effect of their time spent with offspring, and also in light of the possibility that the government wishes to launder the agents' preferences into the social welfare function, different agents would require different adjustments in τ_t. Thus, a single tax instrument, τ_t, has to be tailored in a way that strikes a balance between the adjustment ideally required to correct the behavior of the low-skilled agents and the one ideally required to correct the behavior of the high-skilled agents.

Notice in particular that, since $\zeta_t^j > 0$ for all j but the sign of $\partial \pi^j / \partial n_t^j$ is type-specific, the direction of the required adjustment in τ_t will most likely be the opposite for high- and low-

[11] In section 5 we briefly discuss how results woule be affected if it were possible to tax nonlinearly the purchases of day-care services.

skilled agents. The reason why we do not say that this will certainly be the case is that, if $\varepsilon \neq 1$ so that there is some laundering of the individuals' preferences in the social welfare function, it might happen that the sign of $(1-\varepsilon)\left(H^2 - H^1\right)\eta_t'^j - \gamma_t$ is type-specific too. However, at least for the no-laundering scenario or for small degree of laundering (values of ε which are close to one), we can see that the optimal value of τ_t tends to be pushed up by the concern to affect the allocation of time of high-skilled parents, whereas the concern for affecting the allocation of time of low-skilled parents would call for subsidizing day-care expenditures.

A high value for ζ_t^j reflects a situation where the commodity tax rate is a very effective instrument to alter the demand by agents of type j for day-care services. Because of that, it is also a very effective instrument to affect the amount of time spent by agents of type j with their kids. Thus, the optimal value chosen for τ_t will tend to reflect more strongly how τ_t can be used to indirectly affect the time spent with children by parents of skill type j in the desired direction.

Having characterized the structure of the optimal mixed tax system, we can notice that the so called "principle of targeting" fails to hold. This principle states that a distortion is best addressed by the instrument that acts directly on the relevant margin.[12] In our setting it is the time spent by parents with their kids that represents a source of inefficiency that requires to be corrected by the public policy. Due to the time constraint faced by agents, the time spent by parents with their kids determines the amount of day care services that they demand and vice versa. Thus, one can also equivalently state that it is the amount of day care services used by parents that represents a source of externality that calls for a corrective public intervention. The tax imposed on the purchases of day care services can then be viewed in our setting as the direct instrument to correct the agents' behavior. Therefore, according to the principle of targeting only the tax formula for τ_t should be altered for corrective purposes. But we can easily check that this is not true in our case. In fact, we can see that the terms depending on ε and γ_t do not vanish from the expressions characterizing the marginal tax rates for high- and low-skilled agents even after substituting in (24) and (29) the expression for τ_t provided by (40).

This result is due to the fact that, whereas the required correction to the behavior of high- and low-skilled parents is different, the commodity tax rate τ_t is restricted to be linear and therefore uniform across types. On one hand the government would like to distort in different

[12] See Dixit (1985).

ways the demand for day care services coming from high- and low-skilled parents. On the other hand this would require the possibility to impose a nonlinear tax on the consumption of day care services. However, since agents demand day care services both for time spent working and for leisure time spent without the kids, the government can use the income tax as an additional indirect instrument to affect the agents' demand for day care services. More importantly, the income tax is allowed to be nonlinear and therefore allows the government to let different agents face different marginal tax rates. Under this respect, the nonlinear income tax is designed also to serve the role of a (imperfect) substitute instrument for the impossibility to tax nonlinearly the agents' purchases of day care services.[13]

So far we have been neglecting how the quality of child care services should be chosen.[14] Now suppose that the level of e is optimally chosen period by period. Denoting by q_t the consumer price of child care services, $q_t = p(e_t) + \tau_t$, the first order condition with respect to e_t is given by:

$$\begin{aligned} & \sum_{j=1}^{2}\left[\frac{\partial V_t^j}{\partial e_t} - (1-\varepsilon)\left(H^2 - H^1\right)\eta_t^{\prime j}\left(\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t}\right)\right] f_t^j + \\ & \mu_t \sum_{j=1}^{2} \tau_t \frac{\partial d_t^j}{\partial e_t} f_t^j + \lambda_t \left(\frac{\partial V_t^2}{\partial e_t} - \frac{\partial \widehat{V}_t^2}{\partial e_t}\right) \\ = & -\gamma_t \sum_{j=1}^{2}\left(\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t}\right) f_t^j - p'(e_t)\,\Upsilon_t, \end{aligned}$$

where, similarly to (18), Υ_t has been defined as:

$$\begin{aligned} \Upsilon_t \equiv & \sum_{j=1}^{2}\left[\frac{\partial V_t^j}{\partial q_t} - (1-\varepsilon)\left(H^2 - H^1\right)\eta_t^{\prime j}\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial q_t}\right] f_t^j + \mu_t \sum_{j=1}^{2} \tau_t \frac{\partial d_t^j}{\partial q_t} f_t^j + \lambda_t \left(\frac{\partial V_t^2}{\partial q_t} - \frac{\partial \widehat{V}_t^2}{\partial q_t}\right) + \\ & \gamma_t \sum_{j=1}^{2} \frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial q_t} f_t^j, \end{aligned}$$

[13]In this section we have maintained the assumption that the purchases of day-care services can only be taxed at a constant rate. A possible alternative would be to allow for a fully nonlinear tax system where also the purchases of day-care services are taxed on a nonlinear scale. In terms of results the main difference would be that, under such a system, high-skilled agents would face no distortion with respect to the choice between z and c (i.e., the marginal rate of substitution between z and c would be for them equal to wH^2) whereas the low-skilled agents' choice between z and c, albeit distorted, would solely be distorted for the purpose of deterring mimicking, but not for either externality-correction purposes or for the potential pursuit of non-welfaristic objectives.

[14]Our only assumption has been that the quality level of child care services lies between the level that can be provided at home by low-skilled parents and the one that can be provided at home by high-skilled parents.

and $p'(e_t)\Upsilon_t$ captures the effects of the increase in the unitary price of child care services (due to the higher quality level) on the Lagrangian of the government's problem.

Now define by $MRS_{ec}^{j,t}$ the marginal rate of substitution between the quality of child care services and private consumption (for a given value of q_t) for an agent of type j at time t. Thus, we have:

$$MRS_{ec}^{j,t} = \frac{\partial V_t^j}{\partial e_t} / \frac{\partial V_t^j}{\partial B_t^j} = \left(H^2 - H^1\right) \eta_t^{\prime j} \frac{\partial \pi^j}{\partial e_t} / \frac{\partial V_t^j}{\partial B_t^j}. \tag{41}$$

Adding and subtracting $\lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \frac{\partial V_t^1}{\partial e_t} / \frac{\partial V_t^1}{\partial B_t^1}$ and rearranging terms allow to rewrite the first order condition with respect to e_t as:

$$\begin{aligned}
& \left(f_t^1 \frac{\partial V_t^1}{\partial B_t^1} - \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1}\right) \frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}} + \left(f_t^2 + \lambda_t\right) \frac{\partial V_t^2}{\partial B_t^2} \frac{\frac{\partial V_t^2}{\partial e_t}}{\frac{\partial V_t^2}{\partial B_t^j}} + \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \left(\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}} - \frac{\frac{\partial \widehat{V}_t^2}{\partial e_t}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} \right) + \\
& \sum_{j=1}^{2} \left[-\left(1-\varepsilon\right)\left(H^2 - H^1\right) \eta_t^{\prime j} \left(\frac{\partial \pi^j}{\partial n_t^j} \frac{\partial n_t^j}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t} \right) \right] f_t^j + \mu_t \sum_{j=1}^{2} \tau_t \frac{\partial d_t^j}{\partial e_t} f_t^j \\
= \; & -\gamma_t \sum_{j=1}^{2} \left(\frac{\partial \pi^j}{\partial n_t^j} \frac{\partial n_t^j}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t} \right) f_t^j - p'(e_t)\Upsilon_t.
\end{aligned} \tag{42}$$

Now use (22) and (27) to get expressions for respectively $\left(f_t^2 + \lambda_t\right) \frac{\partial V_t^2}{\partial B_t^2}$ and $f_t^1 \frac{\partial V_t^1}{\partial B_t^1} - \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1}$ and substitute in (42). This gives:

$$\begin{aligned}
& \left\{ \left[(1-\varepsilon)\left(H^2 - H^1\right)\eta_t^{\prime 1} - \gamma_t \right] \frac{\partial \pi^1}{\partial n_t^1} \frac{\partial n_t^1}{\partial B_t^1} - \mu_t \left(-1 + \tau_t \frac{\partial d_t^1}{\partial B_t^1} \right) \right\} f_t^1 \frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}} + \\
& \left\{ \left[(1-\varepsilon)\left(H^2 - H^1\right)\eta_t^{\prime 2} - \gamma_t \right] \frac{\partial \pi^2}{\partial n_t^2} \frac{\partial n_t^2}{\partial B_t^2} - \mu_t \left(-1 + \tau_t \frac{\partial d_t^2}{\partial B_t^2} \right) \right\} f_t^2 \frac{\frac{\partial V_t^2}{\partial e_t}}{\frac{\partial V_t^2}{\partial B_t^j}} + \mu_t \sum_{j=1}^{2} \tau_t \frac{\partial d_t^j}{\partial e_t} f_t^j + \\
& \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \left(\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}} - \frac{\frac{\partial \widehat{V}_t^2}{\partial e_t}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} \right) - \sum_{j=1}^{2} \left[(1-\varepsilon)\left(H^2 - H^1\right) \left(\frac{\partial \pi^j}{\partial n_t^j} \frac{\partial n_t^j}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t} \right) \eta_t^{\prime j} \right] f_t^j \\
= \; & -\gamma_t \sum_{j=1}^{2} f_t^j \left(\frac{\partial \pi^j}{\partial n_t^j} \frac{\partial n_t^j}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t} \right) - p'(e_t)\Upsilon_t.
\end{aligned} \tag{43}$$

From (33) we can easily see that at an optimum $\Upsilon_t = -\mu_t \sum_{j=1}^{2} d_t^j f_t^j$. Therefore, dividing by

μ_t all terms in the previous equation and rearranging terms, we get:

$$\sum_{j=1}^{2}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}f_t^j=-\frac{\gamma_t}{\mu_t}\sum_{j=1}^{2}\left(\frac{\partial\pi^j}{\partial e_t}+\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial e_t}-\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial B_t^j}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}\right)f_t^j+$$

$$\tau_t\sum_{j=1}^{2}\left(\frac{\partial d_t^j}{\partial B_t^j}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}-\frac{\partial d_t^j}{\partial e_t}\right)f_t^j-\frac{\lambda_t}{\mu_t}\frac{\partial\widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}}-\frac{\frac{\partial\widehat{V}_t^2}{\partial e_t}}{\frac{\partial\widehat{V}_t^2}{\partial B_t^1}}\right)+p'(e_t)\sum_{j=1}^{2}d_t^jf_t^j+$$

$$+\frac{1}{\mu_t}\sum_{j=1}^{2}(1-\varepsilon)\left(H^2-H^1\right)\left(\frac{\partial\pi^j}{\partial e_t}+\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial e_t}-\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial B_t^j}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}\right)f_t^j\eta_t'^j.$$

Defining $\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0}$ as $\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0}\equiv\frac{\partial\pi^j}{\partial e_t}+\left(\frac{\partial n_t^j}{\partial e_t}-\frac{\partial n_t^j}{\partial B_t^j}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}\right)\frac{\partial\pi^j}{\partial n_t^j}$ and $\left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0}$ as $\left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0}\equiv\frac{\partial d_t^j}{\partial e_t}-\frac{\partial d_t^j}{\partial B_t^j}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}$, we can express the condition implicitly defining the optimal level of child care quality as:[15]

$$\sum_{j=1}^{2}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}f_t^j=\frac{1}{\mu_t}\sum_{j=1}^{2}\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^j-\gamma_t\right]\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0}f_t^j-\tau_t\sum_{j=1}^{2}\left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0}f_t^j+$$

$$\frac{\lambda_t}{\mu_t}\frac{\partial\widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial\widehat{V}_t^2}{\partial e_t}}{\frac{\partial\widehat{V}_t^2}{\partial B_t^1}}-\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}}\right)+p'(e_t)\sum_{j=1}^{2}d_t^jf_t^j.$$

Using (41) the equation above can be rewritten as:

$$\sum_{j=1}^{2}f_t^jMRS_{ec}^{j,t}=p'(e_t)\sum_{j=1}^{2}d_t^jf_t^j+\frac{1}{\mu_t}\sum_{j=1}^{2}\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^j-\gamma_t\right]\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0}f_t^j+$$

$$\frac{\lambda_t}{\mu_t}\frac{\partial\widehat{V}_t^2}{\partial B_t^1}\left(\widehat{MRS}_{ec}^{2,t}-MRS_{ec}^{1,t}\right)-\tau_t\sum_{j=1}^{2}\left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0}f_t^j. \tag{44}$$

Eq. (44) can be interpreted as a sort of modified Samuelson-type condition, although it does not refer to the efficient level of provision of a public good. The term on the left hand side of eq. (44) measures the sum of the agents' marginal willingness to pay for an increased level of quality of child-care services. The first term on the right hand side of (44) represents the additional resource cost of raising the quality of child-care services, keeping fixed the consumption of services by agents. It is the only term that would be left in a setting

[15] Notice that we can here disregard the effects of the change in the consumer price of child-care services since they are taken care of by the Υ-term.

where: i) asymmetric information problems were absent (no self-selection constraints in the government's problem); ii) the government's objective function were welfaristic, which means that there is no laundering ($\varepsilon = 1$); iii) externalities were absent, in the sense that there is no need to correct agents' behavior at period t to induce them to internalize the social value of increasing the proportion of high-skilled agents at period $t+1$. Discounting for the fact that we are forcing agents to consume the same quality level of child care services, we could regard the condition $\sum_{j=1}^{2} f_t^j MRS_{ec}^{j,t} = p'(e_t) \sum_{j=1}^{2} d_t^j f_t^j$ as a first-best benchmark equating the sum of marginal benefits with the marginal cost of raising quality. Thus, the remaining terms on the right side of (44) describe how an optimizing policy maker should deviate from the first-best rule to take into account self-selection problems, non-welfaristic objective functions and externalities. One can notice that the presence of the last term on the right side of (44) does not challenge this interpretation because, as evident from (40), a commodity tax/subsidy on day-care services can only be justified based on self-selection problems, non-welfaristic objective functions or externalities.

The second term on the right side of (44) reflects how the possibility to vary the quality of day care services can be used for externality-correction purposes and for the potential pursuit of non-welfaristic objectives. An increase in the quality of child care services exerts both a direct and an indirect effect on the probability that the child of a type j parent becomes a high-skilled adult. The direct effect is due to the fact that the quality of child care services enters as an argument into the function π^j. The indirect effect is due to the fact that a change in the quality level will in general induce parents to modify their decisions on the allocation of time. Both these effects are captured by $\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0}$, which also captures how parents vary the amount of time spent with their children in response to a variation in disposable income intended to leave their utility unchanged when the level of quality is marginally increased. The sign of $\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0}$ is therefore in general ambiguous. However, if we make the assumption that the direct effect of an increase in the quality level dominates the indirect effects, we will have that $\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0} > 0$. Assuming moreover that the degree of laundering of agents' preferences in the government's objective function is nil or close to zero ($\varepsilon \to 1$), we can conclude that the sign of the second term on the right side of (44) is negative, which pushes for an increase in the second-best efficient level of child care quality.

The third term on the right side of (44) is a self-selection term that depends on the difference between a mimicker's marginal willingness to pay for increased child-care quality and the corresponding marginal willingness to pay for a true low-skilled. If we assume that, having more time to devote to non-market activities, a mimicker spends more time with his child and therefore spends less for day-care services, the marginal utility of consumption tends to be lower for a mimicker than for a true low-skilled. Taking this into account, (41) tends to imply that the marginal willingness to pay for increased quality is larger for a mimicker than for a true low-skilled.[16] In terms of the effects on the rule governing the optimal level of quality of day-care services, this can be interpreted as an increase in the net marginal cost of raising quality. The underlying intuition is that, as the mimicker's marginal willingness to pay for quality is larger, a marginal increase in quality, accompanied by a change in the income tax payment of the low-skilled agent that leaves his utility unaffected, would make a mimicker better off and thus would tighten the self-selection constraint.

Finally, the last term on the right side of (44) provides an account of how government's (commodity tax) revenues are affected by a change in the agents' consumption pattern when a compensated increase in child-care quality is implemented. Assuming that agents' consumption of day care services go up when the quality of services increases, the last term on the right side of (44) raises (resp.: lowers) the net marginal cost of quality whenever the purchases of day-care services is subsidized (resp.: taxed at a positive rate) by the government.

5 Extensions

In this section we briefly consider two extensions of the model that we have analyzed in the previous section.

The first extension that we consider is obtained assuming that some parents have a faulty belief about the shape of the function that relates the amount of time they spend with their offspring with the probability that the offspring will become high-skilled adults. For illustrative purposes, we consider here the case where low-skilled agents have wrong beliefs. Formally, this means that they take decisions based on the function $\overline{\pi}^1(n_t^1)$ whereas $\pi^1(n_t^1)$ is the true function relating n_t^1 to the probability the child will be of a high-skilled type. We also assume that low-skilled agents tend to underestimate the negative effect that n_t^1 is going to have on the

[16] It is however clear from (41) that one should also consider how the numerator of the expression defining the marginal willingness to pay for quality differs for a mimicker and a true low-skilled. In our discussion here we are for simplicity disregarding the possibility that this effect more than offsets the effect that works through the difference in the denominators.

expected human capital of their children. Under the aforementioned assumptions, the design problem of the government is summarized by the following Lagrangian:

$$\begin{aligned}
\pounds \;=\; & \sum_{t=0}^{\infty}\rho^t \left[V_t^1 + \varepsilon\eta\left(\pi^1\left(n_t^1\right)H^2 + \left(1-\pi^1\left(n_t^1\right)\right)H^1\right) - \eta\left(\overline{\pi}^1\left(n_t^1\right)H^2 + \left(1-\overline{\pi}^1\left(n_t^1\right)\right)H^1\right)\right] f_t^1 + \\
& \sum_{t=0}^{\infty}\rho^t \left[V_t^2 - (1-\varepsilon)\eta\left(\pi^2\left(n_t^2\right)H^2 + \left(1-\pi^2\left(n_t^2\right)\right)H^1\right)\right] f_t^2 + \\
& \sum_{t=0}^{\infty}\rho^t \mu_t \sum_{j=1}^{2}\left(Y_t^j - B_t^j + \tau_t d_t^j\right) f_t^j - \sum_{t=0}^{\infty}\rho^t\gamma_t\left[f_{t+1}^2 - \sum_{j=1}^{2} f_t^j \pi^j\left(n_t^j, e\right)\right] + \\
& \sum_{t=0}^{\infty}\rho^t\lambda_t\left(V_t^2 - \widehat{V}_t^2\right).
\end{aligned}$$

To save space, we only consider the effects on the results about the optimal marginal effective tax rates faced by high- and low-skilled agents. With respect to the former, it is quite easy to realize that the result given by (25) is still valid. Intuitively, since we haven't changed any assumption pertaining to the behavior of the high-skilled agents, the structure of the optimal distortion imposed on them should remain unaffected. With respect to the low-skilled agents, instead, the result provided by (30) is no longer valid. If we were to write the new first order conditions with respect to Y_t^1 and B_t^1, and follow a similar procedure to the one that led to expression (30) above, it would be only a matter of tedious calculations to end up with:

$$\begin{aligned}
METR_t^1 \;=\; & \frac{\lambda_t}{\mu_t f_t^1}\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} - \frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\right) + \\
& + \frac{1}{\mu_t}\left[(1-\varepsilon)\left(H^2 - H^1\right)\eta_t'^1\left(EH\right) - \gamma_t\right]\frac{\partial \pi^1}{\partial n_t^1}\left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0} \\
& \underbrace{+\frac{1}{\mu_t}\left(H^2 - H^1\right)\left[\frac{\partial \overline{\pi}^1}{\partial n_t^1}\eta_t'^1\left(\overline{EH}\right) - \frac{\partial \pi^1}{\partial n_t^1}\eta_t'^1\left(EH\right)\right]\left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0}}_{\Gamma},
\end{aligned} \tag{45}$$

where $\overline{EH}$ and EH represent the expected human capital of the child as assessed on the basis of, respectively, the "wrong" function $\overline{\pi}^1(\cdot)$ and the correct function $\pi^1(\cdot)$.

Comparing (45) and (30) we immediately see that the only difference is the presence in the former of the term labelled Γ. This term reflects the difference between the warm-glow effect of a marginal increase in n_t^1, as perceived by low-skilled parents on the basis of the "wrong" function $\overline{\pi}^1(\cdot)$, and the warm-glow effect of a marginal increase in n_t^1 if low-skilled parents were correctly recognizing the shape of the function $\pi^1(\cdot)$. A positive (resp.: negative) sign of Γ tends to imply that the marginal effective tax rate faced by low-skilled agents should be larger (resp.: smaller) when low-skilled agents misperceive the true shape of the function $\pi^1(\cdot)$.

Given our assumption that the low-skilled agents tend to underestimate the negative effect that n_t^1 is going to have on the expected human capital of their children, it is reasonable to assume that $\partial\overline{\pi}^1/\partial n_t^1 > \partial\pi^1/\partial n_t^1$ and $\overline{\pi}^1(n_t^1) > \pi^1(n_t^1)$. The latter inequality implies that $\overline{EH} > EH$, and this in turn implies, due to the assumption that the η-function is increasing and concave, that $\eta_t^{\prime 1}\left(\overline{EH}\right) < \eta_t^{\prime 1}\left(EH\right)$, meaning that low-skilled parents undervalue the true marginal warm-glow effect of increasing n_t^1. Thus, the sign of the term within square brackets in Γ is unambiguously positive, reflecting the circumstance that low-skilled parents undervalue the true (negative) marginal warm-glow effect of increasing n_t^1. Taking into account that $\left(dn_t^1/dY_t^1\right)_{dV_t^1=0} < 0$, we can then conclude that the overall sign of Γ is negative. This result is in accordance with intuition. The fact that low-skilled parents underestimate the negative effect of n_t^1 tends to make them spend too much time with their children, at least too much as compared with the time that would be spent by a utility-maximizing low-skilled parent who correctly perceived the shape of $\pi^1(n_t^1)$. Under such circumstances, a lower marginal effective tax rate represents an instrument to distort the low-skilled agents' behavior in the desired direction, inducing them to raise labor supply and in this way limiting the total amount of time that they can allocate between pure leisure and time with children.

Another possible extension of the model analyzed in section 4 is to introduce type-specific externalities. Suppose for instance that the expected human capital of a child does not only depend on his/her parents' time-allocation decisions but also on those of parents of different types. Formally, let the probability for a parent of type j of having a high-human-capital-child be given by $\pi^j(n_t^j, H^j, e_t, n^{k\neq j})$ rather than by $\pi^j(n_t^j, H^j, e_t)$ as assumed so far. We could then rewrite the government's problem of the section above as:

$$
\begin{aligned}
\pounds \;=\; & \sum_{t=0}^{\infty}\rho^t\sum_{j=1}^{2}\left[V_t^j-(1-\varepsilon)\,\eta\left(\pi^j\left(n_t^j,E_t^j\right)H^2+\left(1-\pi^j\left(n_t^j,E_t^j\right)\right)H^1\right)\right]f_t^j+\\
& \sum_{t=0}^{\infty}\rho^t\mu_t\sum_{j=1}^{2}\left(Y_t^j-B_t^j+\tau_t d_t^j\right)f_t^j-\sum_{t=0}^{\infty}\rho^t\gamma_t\left[f_{t+1}^2-\sum_{j=1}^{2}f_t^j\pi^j\left(n_t^j,E_t^j\right)\right]+\\
& \sum_{t=0}^{\infty}\rho^t\lambda_t\left(V_t^2-\widehat{V}_t^2\right)+\sum_{t=0}^{\infty}\rho^t\left[\theta_t^1\left(E_t^1-n_t^2\right)+\theta_t^2\left(E_t^2-n_t^1\right)\right],
\end{aligned}
$$

where we have denoted by E_t^j the externality faced by agents of type j and where we have added two type-specific externality constraints, with associated multipliers θ_t^1 and θ_t^2, in the Lagrangian of the government.

Suppose for instance that children of low-skilled agents benefit from the interaction with children of high-skilled agents in day-care centers, meaning that a prolonged interaction with children of high-skilled agents increase their probability of becoming high-human-capital adults. Suppose moreover, for the sake of argument, that the reverse is true for children of high-skilled agents, namely that a prolonged interaction with children of low-skilled agents in day-care centers decrease their probability of becoming high-human-capital adults. This would imply that the sign of the multiplier θ_t^1 is positive, since n_t^2 generates a negative externality on low-skilled agents, whereas the sign of the multiplier θ_t^2 is negative, since n_t^1 generates a positive externality on high-skilled agents.

What would this imply for the conclusions of our model? Also here, to save space, let's focus on the effects on the marginal effective tax rates faced by the different groups of agents. Solving the model again, we would obtain that the marginal effective tax rates faced by high- and low-skilled agents would be respectively given by:

$$
METR_t^2=\frac{1}{\mu_t}\left\{\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^2-\gamma_t\right]\frac{\partial\pi^2}{\partial n_t^2}+\frac{\theta_t^1}{f_t^2}\right\}\left(\frac{dn_t^2}{dY_t^2}\right)_{dV_t^2=0}; \tag{46}
$$

$$
\begin{aligned}
METR_t^1 \;=\; & \frac{\lambda_t}{\mu_t f_t^1}\frac{\partial\widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial\widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial\widehat{V}_t^2}{\partial B_t^1}}-\frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\right)+\\
& \frac{1}{\mu_t}\left\{\left[(1-\varepsilon)\left(H^2-H^1\right)\eta_t'^1-\gamma_t\right]\frac{\partial\pi^1}{\partial n_t^1}+\frac{\theta_t^2}{f_t^1}\right\}\left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0}.
\end{aligned} \tag{47}
$$

From a formal point of view, the only difference with respect to the corresponding formulas of the section above (eqs. (25) and (30)) is the presence in (46) and (47) of a new Pigouvian term. Given our assumption that $\theta_t^2 < 0 < \theta_t^1$, the sign of the new Pigouvian term is, in both (46) and (47), opposite to the sign of the γ-term, pushing in the direction of lowering the marginal effective tax rate faced by the high-skilled agents and raising the marginal effective tax rate faced by the low-skilled agents. The first effect is explained by noticing that a reduction in the marginal effective tax rate faced by the high-skilled agents is a way to induce them to internalize the negative externality that they impose on low-skilled agents by spending more time with their own kids: a lower marginal effective tax rate on high-skilled agents provides them with an incentive to work more and thus let their kids spend more time in day-care centers. A similar reasoning explains why the new Pigouvian term appearing in (47) calls for an increase in the marginal effective tax rate faced by low-skilled agents.

6 Numerical analysis (PRELIMINARY VERSION)

We do not perform a detailed calibration on a specific country. The purpose of the numerical analysis is to illustrate the working of the model and to get a preliminary idea of the relevance of the mechanisms studied in the theoretical part: how strong should the link between child care arrangements and child development be, to play a role in the design of the optimal policies? Given the available estimates, should we care about the effect of child care on child development when we identify the optimal fiscal policy of the government?

To answer this questions, we compare two models:

- Model 1: Day care arrangements have an impact on the distribution of skills (This is the model studied in the previous sections).
- Model 2: "Standard" model where day care arrangements do not have such an impact.

More precisely, we proceed in the following way:

- We assume specific parametric functional forms for the utility function, for the probability of becoming high skilled and for the relationship between the cost of day care services and their quality.
- Both for Model 1 and Model 2, we found the parameters characterizing these functions using a calibration procedure.

- Once these parameters have been found we compute the optimal policy chosen by government both in Model 1 and in Model 2.

- In order to assess the relevance of introducing the link between child care arrangements and child development into an optimal taxation framework, we compute the welfare loss associated with the use in Model 1 of the optimal policies computed in Model 2.

We focus on the model described in Section 2 (that is we don't study take into account the extensions proposed in Section 5) and we assume

- $\epsilon = 1$

- perfect correlation between market ability and ability to rear children; as a consequence we have two types of agents: agents of type 1 are people with no college degree and agents of type 2 are people with a college degree.

As to the optimal choice of the income tax, we assume that the government has to choose within the following class of parametric tax functions (see Li and Sarte 2004):

$$T^j = \zeta * (Y^j)^{1+\phi} \tag{48}$$

with $Y^j = wH^j l^j$. Thus the marginal tax rate is:

$$T'^j = (1+\phi) * \zeta * (Y^j)^{\phi} \tag{49}$$

and the average tax rate is:

$$\bar{T}^j = \zeta * (Y^j)^{\phi} \tag{50}$$

The government can also uses a lump sum Ψ_t and the tax on day care τ_d. When $\phi = 0$ we have the fully linear case of Section 3. As explained below, a tax on consumption is used when the model is calibrated but it is normalized to 0 in the computation of the optimal policies.

We restrict our attention to the steady state.

6.1 Functional forms

We assume the following functional form for the utility function

$$U^j = \alpha \log c^j + \beta_j \log z^j + \gamma_j \log n^j + \delta \log \left(\pi^j(n^j)H^2 + (1-\pi^j(n^j))H^1\right) \tag{51}$$

with $\alpha + \beta_j + \gamma_j + \delta = 1$.

The probability of becoming high skilled is:

$$\pi^j(n^j) = \iota^j + \sigma \frac{x^j}{1 + x^j} \tag{52}$$

where x^j is child care received by an agent of type j and it is produced combining parental time and day care services using the following CES function

$$x^j = \left[\left(n^j H^j \right)^{\frac{1}{\nu}} + \left(d^j e \right)^{\frac{1}{\nu}} \right]^{\nu} \tag{53}$$

with $d^j = \bar{a} - n^j$.

The relationship between the cost p of day care services and their quality e is specified as follows

$$p(e) = \left(\frac{e}{\rho} \right)^{\frac{1}{\kappa}} \tag{54}$$

where $1/\kappa$ is the elasticity of p with respect to e (and κ is the elasticity of e with respect to p). The higher is the quality of day care services the higher is their cost.

6.2 Parametrization and Calibration

We set $\bar{a} = 0.2$ (five years of child care over a period of 25 years). In equation (4) we normalize A to 1.

Parametrization and Calibration: Fiscal variables

As to the tax function (48) we calibrate ζ to have an average tax burden ($\bar{T} = (T^1 + T^2)/(Y^1 + Y^2)$) equal to 35% (see McDaniel 2007 tax data series) and we choose $\phi = 0.5$. We set $\tau_d = -70\%$. Finally, since we observe in the data a positive tax on consumption, we set $\tau_c = 25\%$ (see McDaniel 2007 tax data series).

Parametrization and Calibration: Allocation of time

We normalize to 1 the market productivity of agents of type 1: $H^1 = 1$. The market productivity of agents of type 2 H^2 and the parameters in the utility function $(\alpha, \beta_j, \gamma_j, \delta)$ are chosen in order to generate a realistic allocation of time between labor, child care and leisure for the two groups of agents. For this purpose we consider average data coming from the Harmonized European Time Use Survey (HETUS).[17] Assuming, as it is usually done

[17] The countries we consider are: Finland, France, Italy, Norway, Sweden, United Kingdom. Data refer to people in the age group 25-50.The period considered is 1999-2004.

(e.g. Ragan 2006, Cardia and Ng 2003, Juster 1985), that non-personal time available for discretionary use amounts to 100 hours per week, we have: $l_1 = 30\% < l_2 = 35\%$, $n_1 = 6\% < n_2 = 7\%$, $z_1 = 64\% > z_2 = 58\%$.[18] It should be stressed that high skilled agents devote more time to their kids, though they also work more (see also Guryan, Hurst, and Kearney 2008). Notice that:

- In Model 1 the value of δ affects the allocation of time since the n has an impact on the utility function through π^j. Thus we don't need to assume heterogeneous preferences to match the allocation of time of the two groups; thus we assume $\beta_j = \beta$ and $\gamma_j = \gamma$.

- In Model 2 δ does not affect the allocation of time since π^j does not depend on n (We normalize $\delta = 0$). Thus to match the allocation of time of the two groups we need to allow for different β_j and γ_j. More generally, assuming heterogeneous preferences is the only way to to get $n_1 < n_2$ and $l_1 < l_2$ in Model 2.

Parametrization and Calibration: Child care and human capital

The parameters of the probability function (52) (ι^1, ι^2, σ) and the quality of day care services (e) are chosen to match the following target variables:

- Probabilities of becoming an high skilled: $\pi^1 = 0.23$, $\pi^2 = 0.65$ (see Caucutt, Imrohoroglu, and Kumar 2003), which imply $f^1 = 0.6$ $f^2 = 0.4$.

- $\pi_n^1(n_t)/\pi^1(n_t) < 0$ and $\pi_n^2(n_t)/\pi^2(n_t) > 0$. Thus the effects of parental time on the probability to become skilled is positive (negative) for skilled (unskilled) parents. More precisely we first set $\pi_n^1(n_t)/\pi^1(n_t) = -4\%$ and $\pi_n^2(n_t)/\pi^2(n_t) = 4\%$; we then perform a sensitivity analysis using $\pi_n^1(n_t)/\pi^1(n_t) = -2\%$ and $\pi_n^2(n_t)/\pi^2(n_t) = 2\%$

The parameter ν which affects the elasticity of substitution between parental care and non parental care (i.e. $1/(1-\nu)$) is set equal to 0.9.

[18] Two remarks are important in interpreting these data. First, child care is simply defined as the sum of the minutes registered as devoted to primary and secondary child care: this amount of time is lower than the total time spent with children and it better captures deliberate child care by parents which is the focus of this paper. Second, leisure is here defined as a residual category, that is, it is the time not spent either working or doing primary and secondary child care: as a consequence, it is not a measure of pure leisure as it also includes, for instance, housework.

As to the elasticity of the quality of day care services with respect to the cost (κ) we perform a sensitivity analysis.

The cost of producing child care p is chosen so that the ratio $p/(wH^1)$ is equal to 50% (this is consistent with the evidence in Blomquist, Christiansen, and Micheletto 2010)

6.3 Simulation: optimal policy and welfare evaluation

Once we found the parameters of equations (51)-(54) using the procedure described above, we compute the optimal policy. The government chooses a tax rate τ_d on day care and a specific tax function within the class of parametric functions described in equation (48). The lump sum transfer Ψ_t is set to keep the government budget balanced. Notice that τ_c can be normalized to 0.

Tables 1 and 2 report the results under two different assumptions on the values of $\frac{\pi_n^j(n_t)}{\pi^j(n_t)}$ used in the calibration procedure. In these tables we compute the optimal values of the quality of day care service e, of the parameter ϕ (which affects the degree of progressivity), of the average tax burden $\bar{T} = (T^1 + T^2)/(Y^1 + Y^2)$ and of the tax rate on day care services τ_d. We also show the values of $\frac{\pi_n^j(n_t)}{\pi^j(n_t)}$ computed at the optimum. The first column concerns Model 2 in which day care arrangements do not affect by assumption the probability of becoming skilled ($\sigma = 0$ and thus $\frac{\pi_n^j(n_t)}{\pi^j(n_t)} = 0$); in this case the optimal problem of the government does not obviously concern the choice of the quality of day care services. The other columns shows the result for Model 1, depending on the assumed value of κ which is the elasticity of the quality of day care services with respect to their cost. To interpret these results it may be useful to remark that the values of e and H_2 determined according to the calibration procedure described above are respectively equal to 1.6 and 3.3.

In order to assess the relevance of introducing the link between child care arrangements and child development into an optimal taxation framework, we compute the welfare loss associated with the use in Model 1 of the optimal policies computed in Model 2 instead of the correct ones. We measure this welfare loss as the percentage increase in consumption of both types of agents needed to leave social welfare unchanged. The results are reported in the last row of Tables 1 and 2.

The higher κ is, the higher is the welfare loss due to the computation of the optimal policies using Model 2, though the true model is Model 1.

Table 1: Optimal policies. Assumptions: $\frac{\pi_n^1(n_t)}{\pi^1(n_t)} = -4\%$ $\frac{\pi_n^2(n_t)}{\pi^2(n_t)} = 4\%$

	Model 2	Model 1: $\kappa = 0.4$	Model 1: $\kappa = 0.5$	Model 1: $\kappa = 0.6$	Model 1: $\kappa = 0.7$
e	-	2.1	2.7	3.6	5
ϕ	0	0	0	0	0
$\bar{T}$	35%	35%	40%	40%	40%
τ_d	-70%	-90%	-95%	-95%	-95%
$\frac{\pi_n^1(n_t)}{\pi^1(n_t)}$	0	-5.7%	-7.1%	-8.1%	-8.4%
$\frac{\pi_n^2(n_t)}{\pi^2(n_t)}$	0	3.1%	1.7%	0.02%	-1.1%
Δcons.		2.2%	5.9%	11.7%	19.8%

Table 2: Optimal policies. Assumptions: $\frac{\pi_n^1(n_t)}{\pi^1(n_t)} = -2\%$ $\frac{\pi_n^2(n_t)}{\pi^2(n_t)} = 2\%$

	Model 2	Model 1: $\kappa = 0.4$	Model 1: $\kappa = 0.5$	Model 1: $\kappa = 0.6$	Model 1: $\kappa = 0.7$
e	-	1.7	2.	2.5	3.2
ϕ	0	0	0	0	0
$\bar{T}$	35%	35%	35%	35%	40%
τ_d	-70%	-80%	-80%	-80%	-85%
$\frac{\pi_n^1(n_t)}{\pi^1(n_t)}$	0	$-2.\%$	-2.9%	-3.9%	-4.7%
$\frac{\pi_n^2(n_t)}{\pi^2(n_t)}$	0	2.%	1.7%	1%	-0.4%
Δcons.		0%	0.66%	2.%	4.5%

7 Summary and Discussion

The paper characterizes the optimal tax policy and the optimal level of quality of day care in a two-type OLG model where parental choices over child care determine the probability of having a high skill child in a type-specific way. We consider two different scenarios: first, one where the government can use linear taxation on labor income and a linear tax/subsidy on day care. Second, a set-up where the government can resort to nonlinear taxation of labor income and again a linear tax/subsidy on day care. In both cases we discuss the rules dictating the optimal choice of day care quality enforced by the government.

With respect to previous contributions, optimal tax formulas incorporate two new sets of terms. The first depends on the extent to which the social welfare function reflects the warm-glow component of parental preferences. The second depends on the social marginal utility of turning an unskilled individual into a skilled one.

We also find that the so called "principle of targeting" fails to hold in our setting. It is not just the formula characterizing the tax rate on day care purchases that is affected for externality-correcting purposes; all tax formulas are affected.

Throughout the paper we have assumed that the government's objective function is given by the discounted sum of (adjusted) utilities; the only deviation from a purely utilitarian setting was given by the possibility to launder, fully or partially, the warm-glow component of an agent's utility in the social welfare function. An alternative could have been to consider a social welfare function where the utility of each type of agents is weighted by a given welfare weight rather than by the proportion of that type of agents in the population. Suppose for instance that the government assigns a welfare weight α^i to agents of type i. Then, the only effect on the tax- and quality-of-child-care-formulas presented in the previous sections would have been to multiply by α^j all the terms $(1-\varepsilon)\left(H^2 - H^1\right)\eta_t^{\prime j}$, wherever they appear. In other words, differentiated welfare weights for different types of agents would be equivalent, in terms of the structure of our formulas, to impose type-specific laundering parameters ε. In particular, a relatively large (resp.: small) welfare weight for a given type of agent would have similar effects of choosing a relatively small (resp.: large) value of ε for that type of agent in the social welfare function.

An interesting extension concerns the use of a two sector model to allow for a more detailed study of the production of day care. This issue is on our research agenda.

References

Andreoni, James. 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy* 97 (6): 1447–58 (December).

———. 2006. Chapter 18 of *Philanthropy*, edited by S. Kolm and Jean Mercier Ythier, Volume 1 of *Handbook on the Economics of Giving, Reciprocity and Altruism*, 1201–1269. Elsevier.

Atkinson, A. B., and J. E. Stiglitz. 1976. "The design of tax structure: Direct versus indirect taxation." *Journal of Public Economics* 6 (1-2): 55–75.

Balestrino, Alessandro. 2000. "Mixed Tax Systems and the Public Provision of Private Goods." *International Tax and Public Finance* 7 (4): 463–478 (August).

Bernal, R. 2008. "The effect of maternal employment and child care on children's cognitive development." *International Economic Review*. (forthcoming).

Bernal, R., and M. Keane. 2007. "Child care choices and children's cognitive achievement: the case of single mothers." mimeo.

———. 2008. "Quasi-Structural Estimation of a Model of Child-Care Choices and Child Cognitive Ability Production." *Journal of Econometrics*. (forthcoming).

Blomquist, Sren, Vidar Christiansen, and Luca Micheletto. 2010. "Public Provision of Private Goods and Nondistortionary Marginal Tax Rates." *American Economic Journal: Economic Policy*. forthcoming.

Blomquist, Sren, and Luca Micheletto. 2009. "Nonlinear Income Taxation And Matching Grants In A Federation With Decentralized In-Kind Transfers." *International Economic Review* 50 (2): 543–575 (05).

Cardia, E., and S. Ng. 2003. "Intergenerational time transfers and childcare." *Review of Economic Dynamics* 6:431–454.

Carneiro, P., and J. Heckman. 2002. "The evidence on credit constraints in post-secondary schooling." *Economic Journal* 112:705–734.

———. 2003. "Human capital policy." In *Inequality in America: What Role for Human Capital Policies?*, edited by J Heckman and A Krueger. MIT Press.

Carneiro, Pedro, Costas Meghir, and Matthias Parey. 2007, September. "Maternal Education, Home Environments and the Development of Children and Adolescents." Iza discussion papers 3072, Institute for the Study of Labor (IZA).

Caucutt, Elizabeth M., Selahattin Imrohoroglu, and Krishna B. Kumar. 2003. "Growth and Welfare Analysis of Tax Progressivity in a Heterogeneous-Agent Model." *Review of Economic Dynamics* 6 (3): 546–577 (July).

Cremer, H., and P. Pestieau. 2006a. "Intergenerational Transfer of Human Capital and Optimal Education Policy." *Journal of Public Economic Theory* 8:529–545.

Cremer, Helmuth, and Pierre Pestieau. 2006b. "Intergenerational Transfer of Human Capital and Optimal Education Policy." *Journal of Public Economic Theory* 8 (4): 529–545 (October).

Cuhna, F., J. Heckman, L. Lochner, and D. Masterov. 2005. "Interpreting the evidence on life cycle skill formation." In *Handbook of the Economics of Education*, edited by E. Hanushek and F. Welch. North Holland.

Cunha, Flavio, and James Heckman. 2007. "The technology of skill formation." *American Economic Review* 97:31–47.

Diamond, Peter. 2006. "Optimal tax treatment of private contributions for public goods with and without warm glow preferences." *Journal of Public Economics* 90 (4-5): 897–919 (May).

Dixit, Avinash. 1985. "Tax policy in open economies." Chapter 6 of *Handbook of Public Economics*, edited by A. J. Auerbach and M. Feldstein, Volume 1 of *Handbook of Public Economics*, 313–374. Elsevier.

Edwards, J., M. Keen, and M. Tuomala. 1994. "Income tax, commodity taxes and public good provision: a brief guide." *FinanzArchiv* 51:472–487.

Galor, O., and J. Zeira. 1993. "Income Distribution and Macroeconomics." *Review of Economics Studies* 60:35–52.

Glomm, G., and M. Kaganovich. 2003. "Distributional effect of public education in an economy with public pensions." *International Economic Review* 44:917–937.

———. 2008. "Social security, public education and the growth inequality relationship." *European Economic Review* 52:1009–1034.

Glomm, G., and B. Ravikumar. 1992. “Public versus private investment in human capital: endogenous growth and income inequality.” *Journal of Political Economy* 100:818–834.

Guryan, Jonathan, Erik Hurst, and Melissa Kearney. 2008. “Parental Education and Parental Time with Children.” *Journal of Economic Perspectives* 22 (3): 23–46 (Summer).

Hammond, P. 1987. Chapter Altruism of *The New Palgrave: A Dictionary of Economics* . London: Macmillan.

Heckman, J., and D. Masterov. 2007. “The productivity argument for investing in young children.” mimeo.

Juster, F.T. 1985. “A note on recent changes in time use.” In *Time, Goods, and Well-Being*, edited by F.T Juster and F.P Stafford. Institute for Social Research, University of Michigan.

Kaplow, Louis. 1995. “A note on subsidizing gifts.” *Journal of Public Economics* 58 (3): 469–477 (November).

———. 1998. “Tax Policy and Gifts.” *American Economic Review* 88 (2): 283–88 (May).

Kimmel, Jean, and Rachel Connelly. 2007. “Mothers’ time choices: Caregiving, Leisure, Home Production, and Paid Work.” *Journal of Human Resources* 42, no. 3.

Li, W, and P-D Sarte. 2004. “Progressive Taxation and Long-Run Growth.” *American Economic Review* 94 (5):1705–1716.

McDaniel, C. 2007. “Average tax rates on consumption, investment, labor and capital in the OECD 1950-2003.”

Ragan, Kelly. 2006. “Taxes, transfers and time use: fiscal policy in a household production model.” mimeo.

Ruhm, C.J. 2004. “Parental employment and child cognitive development.” *Journal of Human Resources* 39:155–192.

Schokkaert, E. 2006. Chapter 2 (”The empirical analysis of transfer motives”) of *Handbook of the economics of giving, altruism and reciprocity*, edited by S-C. Kolm and J. M. Ythier, 128–181. North-Holland.

Stiglitz, Joseph E. 1982. “Self-selection and Pareto efficient taxation.” *Journal of Public Economics* 17 (2): 213–240 (March).