Società italiana di economia pubblica
XXI CONFERENZA

PUBLIC CHOICE E POLITICAL ECONOMY
I fondamenti positivi della teoria di finanza pubblica

Pavia, Università, 24-25 settembre 2009

INTERGENERATIONAL TRANSMISSION OF SKILLS DURING CHILDHOOD AND OPTIMAL FISCAL POLICY

ALESSANDRA CASARICO, LUCA MICHELETTO E ALESSANDRO SOMMACAL

società italiana di economia pubblica

dipartimento di economia pubblica e territoriale – università di Pavia

Intergenerational transmission of skills during childhood and optimal fiscal policy*

Alessandra Casarico[†] Luca Micheletto[‡] Alessandro Sommacal[§]

15 June 2009

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 with exogenous growth 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 non linear 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 a two-type OLG model with exogenous growth 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). The features of the model developed are as follows: 1) skills are heterogeneous and the distribution of skills across individuals is endogenous, that is, it is affected by the optimal public policy; 2) the way the parents' time allocation affects the level of human capital of the respective offspring depends on the parents' skills in a non-monotonic way (children of low skilled individuals benefit from public day-care as the quality of day-care centers is higher than the human capital of the parents; the opposite holds for kids of skilled parents. Notice that the quality of child care which is a matter of choice for the government is bounded by the human capital levels in the economy); 3) the model is dynamic: what the parents do influences who the kids are tomorrow through the probability of becoming high-skilled, which is a function of parental child care choices. 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 remark that 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 and analyzed the complementarity between early and late investments in human capital (Bernal, 2008; Bernal and Keane, 2007 and 2008; Carneiro and Heckman, 2003; Cunha and Heckman, 2007; Carneiro, Meghir and Parey, 2007). An explicit inclusion of child care in the skill formation process seems therefore relevant to correctly study the optimal design of the 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. Section 5 offers concluding remarks.

2 The model

2.1 The consumers

We consider a two-period OLG model with intragenerational heterogeneity: agents differ in their skill level; there are only two possible levels of skills H^j, with $H^2 > H^1$. In the first period agents (children) do not take any active choice; depending on child care arrangements and on the human capital of their parents, they have a certain probability to be high- or low-skilled. 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 maximize the following utility function:

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

with $\eta''(\cdot) < 0 < \eta'(\cdot)$, $u''(\cdot) < 0 < u'(\cdot)$, and where c_t^j, z_t^j, n_t^j denote respectively consumption, leisure and time devoted to child care by agent j. The last term in (1) reflects the warm-glow altruism of agents (Andreoni, 1989), who care about the impact that their parental time will have on the probability π^j of having a high-skilled children and therefore on the expected level of human capital of their kids. 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 Banerjee and Newman, 1991; Glomm and Ravikumar 1992; Cremer and Pestieau 2006). 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).

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

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

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

with l_t^j indicating the labour supply 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 first analyze the optimal fiscal policy under the assumption that only linear instruments are available, and we then allow also for non linear taxation.

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 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.[1] We can therefore write the agents' budget constraint in a compact form as:

$$(1+\tau_t^c)c_t^j + (p(e_t)+\tau_t^d)d_t^j = wH^j l_t^j + \Psi_t \tag{4}$$

where the price of consumption is normalized to 1, $p(e_t)$ is the price (net of tax or subsidy τ_t^i) of goods 1 and 2, e_t captures the quality of child care services which is taken as given by the individuals, w is the wage in efficiency units and Ψ_t denotes a lump-sum transfer.

Mixed tax system

A mixed tax system is commonly defined as a system where earned income can be taxed through a nonlinear income tax whereas commodity taxes are allowed to be differentiated across goods but are constrained to be linear. The restriction on the commodity taxes reflects the informational assumption that commodity purchases are observable only at the aggregate level whereas personal consumption of a given good is not observable (or it is observable at a prohibitive cost). Earned income is instead assumed being observable at the individual level and this justifies the possibility to use a nonlinear income tax. Dealing with a model where agents allocate their after-tax income across two goods only, we can choose one of the goods as the untaxed numeraire. We choose for this purpose the consumption good c. Then, denoting by Y_t^j the income earned by an agent of type j at time t, so that $Y_t^j = wH^j l_t^j$, and denoting by $T_t(Y)$ the nonlinear income tax schedule prevailing at time t, the individual's budget constraint is given by:

$$c_t^j = Y_t^j - T_t\left(Y_t^j\right) - \left[p\left(e_t\right)+\tau_t^d\right]d_t^j. \tag{5}$$

2.2 Output

Output is produced according to the following function:

$$Y_t = A(f_t^1 l_t^1 H^1 + f_t^2 l_t^2 H^2) \tag{6}$$

where f^i is the fraction of people of type i. Total population is normalized to 1 and the population growth rate is equal to 0.

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

2.3 Evolution of skills' distribution

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

$$f_{t+1}^2 = \sum_{j=1}^{2} \pi_t^j \cdot f_t^j = \pi_t^2 \cdot f_t^2 + \pi_t^1 \cdot \underbrace{f_t^1}_{1-f_t^2} = \pi_t^2 \cdot f_t^2 + \pi_t^1 \cdot (1 - f_t^2) \tag{7}$$

For the fraction of low skilled we have:

$$f_{t+1}^1 = \sum_{j=1}^{2} (1-\pi_t^j) \cdot f_t^j = (1-\pi_t^2) \cdot \underbrace{f_t^2}_{1-f_t^1} + (1-\pi_t^1) \cdot f_t^1 = (1-\pi_t^2) \cdot (1-f_t^1) + (1-\pi_t^1) \cdot f_t^1 \tag{8}$$

In equation (1) we have used $\pi_t^j = \pi(n_t^j)$, where $\pi^j(n_t^j)$ stands for $\pi(n_t^j, H^j, \bar{e})$, that is the probability of being a type 2 (that is high skilled) agent, is a function of my parents' type j, the time n_t^j they dedicate to child care and the quality of child care services e_t.

We assume that for any given pair (n, e) the following condition holds: $\pi_n(n, H^1, e) < \pi_n(n, H^2, e)$. The underlying assumption is that there is perfect correlation between market ability and ability to raise children (where ability to raise children is meant to capture ability to turn children into high-skilled adults).[2] Moreover, we will also assume that $\pi_n(n_t^1, H^1, e) \leq 0$ and $\pi_n(n_t^2, H^2, e) \geq 0$. This reflect the idea that, keeping labor supply fixed, an additional hour of parental care means a reduction of one hour of day care services; if the quality of the care provided by day care centers reflects the quality mix of the personnel employed at the centers, the "quality" of the care provided is lower (higher) than the "quality" of the care provided by high (lower) skilled agents. If day care centers employ only high-skilled agents or only low-skilled agents, we will not observe any impact of non parental time on the high skilled or on the low skilled.

2.4 Government

As to the government, the objective function is:

$$W = \sum_{t=1}^{\infty} \rho^t \sum_{j=1}^{2} f_t^j \cdot \left\{ u\left(c_t^i, z_t^i, n_t^i\right) + \varepsilon \eta \left(\pi^j(n_t^j) H^2 + (1 - \pi^j(n_t^j)) H^1 \right) \right\} \tag{9}$$

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.

[2] This also means that an averse-to-inequality government will unambiguously try to redistribute from high-skilled agents to low-skilled ones.

Linear tax system

The budget constraints for the government can be written as:

$$\tau_t^c \sum_{j=1}^{2} f_t^j c_t^j + \tau_t^d \sum_{j=1}^{2} f_t^j d_t^j \geq \Psi_t \tag{10}$$

Denoting by $x_t^{jc} = c_t^j$ and $x_t^{jd} = d_t^j$, we can rewrite (10) as:

$$\sum_{i=c,d} \tau_t^i \sum_{j=1}^{2} f_t^j x_t^{ji} \geq \Psi_t \tag{11}$$

The government's budget balances year by year without debt policy.

Mixed tax system

The government budget constraint is:

$$\sum_{j=1}^{2} f_t^j \left(Y_t^j - B_t^j + \tau_t^d d_t^j \right) \geq 0 \tag{12}$$

As above, the government budget is balanced.

3 Linear tax system

3.1 Solution of the consumer optimization problem and indirect utility function

Using the notation introduced above, we rewrite the budget constraint (4) as:

$$(1 + \tau_t^c) x_t^{jc} + (p(e_t) + \tau_t^d) x_t^{jd} = wH^j l_t^j + \Psi_t \tag{13}$$

The maximization of (1) subject to (4) delivers the following first order conditions for the individual problem:

$$\alpha g_t'^j = \chi_t^j (1 + \tau_t^1)$$
$$-\gamma g_t'^j \delta_t'^j - \eta_t'^j \frac{\partial \pi^j}{\partial x_t^{j2}} (H^2 - H^1) = \chi_t^j (p(e_t) + \tau_t^d - H^j)$$
$$\beta g_t'^j v_t'^j = \chi_t^j H^j,$$

where χ_t^j denotes the marginal utility of income for an agent of type j at time t.

3.2 Solution of the government optimization problem

The government maximizes:

$$
\begin{aligned}
\pounds \quad = \quad & \sum_{t=0}^{\infty}\rho^t\sum_{j=1}^{2}\left[V_t^j-(1-\varepsilon)\,\eta\left(\pi^j\left(n_t^j\right)H^2+\left(1-\pi^j\left(n_t^j\right)\right)H^1\right)\right]f_t^j + \\
& \sum_{t=0}^{\infty}\rho^t\mu_t\left(\sum_{i=c,d}\tau_t^i\sum_{j=1}^{2}f_t^jx_t^{ji}-\Psi_t\right)-\sum_{t=0}^{\infty}\rho^t\gamma_t\left[f_{t+1}^2-\sum_{j=1}^{2}f_t^j\pi^j\left(n_t^j\right)\right]
\end{aligned}
\tag{14}
$$

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 $\bar{e}$.

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

$$
\sum_{j=1}^{2}f_t^j\left[\chi_t^j-(1-\varepsilon)\eta_t^{\prime j}\frac{\partial\pi^j}{\partial x_t^{jd}}\frac{\partial x_t^{jd}}{\partial\Psi_t}(H^2-H^1)\right]+\mu_t\left[\sum_{i=c,d}\tau_t^i\sum_{j=1}^{2}f_t^j\frac{\partial x_t^{ji}}{\partial\Psi_t}-1\right]+\gamma_t\sum_{j=1}^{2}f_t^j\frac{\partial\pi^j}{\partial x_t^{jd}}\frac{\partial x_t^{jd}}{\partial\Psi_t}=0.
\tag{15}
$$

which can be rewritten as

$$
E(b_t^j)=1
$$

where $b_t^j=\frac{\chi_t^j}{\mu_t}+\sum_{i=c,d}\tau_t^i\frac{\partial x_t^{ji}}{\partial\Psi_t}-\frac{\partial\pi^j}{\partial x_t^{jd}}\frac{\partial x_t^{jd}}{\partial\Psi_t}\frac{1}{\mu_t}\left[(1-\varepsilon)(H^2-H^1)\eta_t^{\prime j}-\gamma_t\right]$ indicates the net marginal social evaluation of agent j's income. The first term captures the impact that a change in income determined by the lump-sum transfer has on the individual indirect utility function. The second indicates the impact on the government revenues associated with the change in the demand functions of the two goods. The third term shows the impact that a change in the lump-sum transfer has on the demand for child care and therefore on the probability of having a skilled child for agent j. If $\varepsilon=1$, the social evaluation of turning a low skilled into a high skilled is given by γ. When $\varepsilon\neq1$, 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^k. We have:

$$
\begin{aligned}
& \sum_{j=1}^{2}f_t^j\left[-\chi_t^jx_t^{jk}-(1-\varepsilon)\eta_t^{\prime j}\frac{\partial\pi^j}{\partial x_t^{jd}}\frac{\partial x_t^{jd}}{\partial\tau_t^k}(H^2-H^1)\right]+\mu_t\left[\sum_{j=1}^{2}f_t^jx_t^{jk}+\sum_{i=c,d}\tau_t^i\sum_{j=1}^{2}f_t^j\frac{\partial x_t^{ji}}{\partial\tau_t^k}\right]+ \\
& \gamma_t\sum_{j=1}^{2}f_t^j\frac{\partial\pi^j}{\partial x_t^{jd}}\frac{\partial x_t^{jd}}{\partial\tau_t^k}=0
\end{aligned}
\tag{16}
$$

Using the Slutsky equation and denoting hicksian demands by a "tilde", we can rewrite

(16), after rearranging terms, as follows:

$$\frac{\sum_{i=c,d}\tau_t^i\sum_{j=1}^2 f_t^j\partial\widetilde{x}_t^{jk}/\partial\tau_t^i}{x_t^k} = -\left[1-Cov(b_t^j,\frac{x^{jk}}{x^k})\right]+ \frac{1}{\mu_t}\sum_{j=1}^2\frac{\partial\pi^j}{\partial x_t^{jd}}\frac{\partial\widetilde{x}_t^{jd}/\partial\tau_t^k}{x_t^k}f_t^j\left[(1-\varepsilon)(H^2-H^1)\eta_t'^j-\gamma_t\right] \tag{17}$$

The proportional change in the aggregate compensated demand for good k 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 $Cov(b_t^j,\frac{x^{jk}}{x^k})$, the lower should be the reduction of the consumption of good k 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 depends on whether the government takes into account or not the warm-glow component of individual preferences. The second one identifies the externality related to the assumption of imperfect altruism. 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 skill type. For this reason, tax rates have to take care at the same time of the adjustments ideally required to correct the behavior of both skilled and unskilled agents. When $k=d$, that is, when the focus is on day care, the type of adjustment induced by this additional term will most likely be of opposite sign for skilled and unskilled agents. This is certainly the case when $\varepsilon=1$: recalling that $\partial\widetilde{x}_t^{jd}/\partial\tau_t^d<0$, indeed, the sign of $\partial\pi^j/\partial x_t^{jd}$ is positive (negative) if agent j is unskilled (skilled). For an unskilled agent, the presence of this additional positive term is such that, *ceteris paribus*, the higher the term is, the lower should be the reduction of the consumption of child care. The opposite holds for a skilled agent. When $k=c$, the sign of $\partial\widetilde{x}_t^{jc}/\partial\tau_t^d$ depends on whether consumption and day-care are substitutes or complements. If the two goods are complements, that is $\partial\widetilde{x}_t^{jc}/\partial\tau_t^d<0$, all the observations put forward above still apply. If the two goods are substitutes, that is $\partial\widetilde{x}_t^{jc}/\partial\tau_t^d>0$, the term calls for an increase in the consumption of good c for the skilled and for a decrease for the unskilled. When $\varepsilon\neq1$, also the term in the last square brackets is type-specific and we are not guaranteed that the adjustment imposed by this additional term will be of opposite sign for the two types.

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

e_t. Differentiating (14) with respect to e_t, we find:

$$\sum_{j=1}^{2} f_t^j \left[\frac{\partial V_t^j}{\partial e_t} - (1-\varepsilon)\, \eta_t^{\prime j} \left(\frac{\partial \pi^j}{\partial x_t^{jd}} \frac{\partial x_t^{jd}}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t} \right) \left(H^2 - H^1\right) \right] + \tag{18}$$
$$\mu_t \sum_{i=c,d} \tau_t^i \sum_{j=1}^{2} f_t^j \frac{\partial x_t^{ji}}{\partial e_t} + \gamma_t \sum_{j=1}^{2} f_t^j \left(\frac{\partial \pi^j}{\partial x_t^{jd}} \frac{\partial x_t^{jd}}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t} \right) + p'(e_t)\, \Upsilon_t = 0$$

where Υ_t has been defined as:

$$\Upsilon_t \equiv \sum_{j=1}^{2} f_t^j \left[\frac{\partial V_t^j}{\partial q_t} - (1-\varepsilon) \left(H^2 - H^1\right) \eta_t^{\prime j} \frac{\partial \pi^j}{\partial x_t^{jd}} \frac{\partial x_t^{jd}}{\partial q_t} \right] + \mu_t \sum_{i=c,d} \tau_t^i \sum_{j=1}^{2} f_t^j \frac{\partial x_t^{jd}}{\partial q_t} + \tag{19}$$
$$\gamma_t \sum_{j=1}^{2} f_t^j \frac{\partial \pi^j}{\partial x_t^{j2}} \frac{\partial x_t^{jd}}{\partial q_t}.$$

with $(p(e_t) + \tau_t^d) = q_t$. Using the first order condition with respect to τ_t^d (16), it is straightforward to conclude that:

$$\Upsilon_t = -\mu_t x_t^d$$

where $x_t^d = \sum_{j=1}^{2} f_t^j x_t^{jd}$ denotes aggregate consumption of day-care. We can therefore rewrite (18) as:

$$\sum_{j=1}^{2} f_t^j \frac{\partial V_t^j}{\partial e_t} = p'(e_t)\, \mu_t x_t^d - \mu_t \sum_{i=c,d} \tau_t^i \sum_{j=1}^{2} f_t^j \frac{\partial x_t^{ji}}{\partial e_t} + \sum_{j=1}^{2} f_t^j \left[(1-\varepsilon)\, \eta_t^{\prime j} \left(H^2 - H^1\right) - \gamma_t \right] \left(\frac{\partial \pi^j}{\partial x_t^{jd}} \frac{\partial x_t^{jd}}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t} \right).$$

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 $\partial \pi^j / \partial e_t$) and indirectly (the term $\left(\partial \pi^j / \partial x_t^{jd}\right) / \left(\partial x_t^{jd} / \partial e_t\right)$). 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.

Given that the government can observe earned income at an individual level but 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. Here we confine our analysis to 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:

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

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.

Defining $\eta_t'^j$ $(j = 1, 2)$ as $\eta_t'^j = \eta'\left(\pi^j\left(n_t^j\right)H^2 + \left(1-\pi^j\left(n_t^j\right)\right)H^1\right)$, the first order conditions for Y_t^2 and B_t^2 are:

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

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

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 being 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:

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

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

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

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

Since from the optimization problem solved by the high-skilled agents we can implicitly express the marginal tax rate faced by them as $T'\left(Y_t^1\right) = 1 + \left(\frac{\partial V_t^1}{\partial Y_t^1} / \frac{\partial V_t^1}{\partial B_t^1} \right)$, collecting terms in (28) gives:

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

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.[3] 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, 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.

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,[4] it is common practice to make the so-called agent-monotonicity assump-

[3] 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.

[4] See, for example, Stiglitz (1982) and Edwards et al. (1994).

tion. This assumption requires that, at any given point in the (Y, B)-space, the indifference curves are shallower the higher the wage rate of an agent. If this assumption is satisfied, we can safely conclude that the self-selection term takes a positive sign and therefore requires a downwards 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 labor supply but in general also, once expenses on day-care services have been subtracted, with respect to the amount available for private consumption. To explore this issue in more details, take any given bundle in the (Y, B)-space and consider the marginal rate of substitution between Y and B for a generic agent of type i. This is given by $-\left(\partial V^i/\partial Y\right)/\left(\partial V^i/\partial B\right)$. Assuming a utility function of the form $u\left(c^i, z^i, n^i\right) + \eta\left(\pi^i\left(n^i\right)H^2 + \left(1 - \pi^i\left(n^i\right)\right)H^1\right)$, the conditional indirect utility for an agent of type i, $V^i(Y, B)$, is obtained maximizing $u\left(c^i, z^i, n^i\right) + \eta\left(\pi^i\left(n^i\right)H^2 + \left(1 - \pi^i\left(n^i\right)\right)H^1\right)$ subject to the budget constraint $c^i = B - (p(e) + \tau)\, d^i = B - (p(e) + \tau)\left(l^i + z^i\right) = B - (p(e) + \tau)\left[\left(Y/wH^i\right) + z^i\right]$ and the time constraint $n^i = 1 - l^i - z^i$. This implies that $\partial V^i/\partial B = \partial u\left(\cdot; wH^i\right)/\partial c^i$ and $\partial V^i/\partial Y = -\left(wH^i\right)^{-1}\left[\frac{\partial u(\cdot;wH^i)}{\partial c^i}(p(e) + \tau) + \frac{\partial u(\cdot;wH^i)}{\partial n^i} + \left(H^2 - H^1\right)\eta'^i\frac{\partial \pi^i}{\partial n^i}\right]$. Thus, we have:

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

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

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,[5] and given that $d^i = 1 - n^i$, 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 that $\left(H^2 - H^1\right)\widehat{\eta}_t^{\prime 2}\partial \widehat{\pi}^2/\partial \widehat{n}_t^2 > 0$ whereas $\left(H^2 - H^1\right)\eta_t^{\prime 1}\partial \pi^1/\partial n_t^1 < 0$ 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:

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

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:

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

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 $-d_t^1 \frac{\partial V_t^1}{\partial B_t^1}$:

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

[5] 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 et al. (2008).

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

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:

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

Simplifying terms in (37) gives:

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

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:

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

The denominator of the expression on the right hand side of (39) is negative and 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 unaffected the utility of all non-mimicking agents 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 easier the interpretation of the second term appearing at the numerator of (39), it is convenient to introduce the variable ζ_t^j, defined as $\zeta_t^j \equiv \frac{\partial \widetilde{d}_t^j}{\partial \tau_t} f_t^j / \sum_{i=1}^{2} \frac{\partial \widetilde{d}_t^i}{\partial \tau_t} f_t^i$, and rewrite τ_t as

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

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. 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 only taxable linearly, meaning that the commodity tax (or subsidy) rate on day-care purchases is the same for all agents, irrespective of skill type. 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-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 $(1-\varepsilon)\left(H^2 - H^1\right)\eta_t^{\prime j} - \gamma_t$ 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 in the desired direction the time spent with children by parents of skill type j.

So far we have been neglecting how the quality of child care services should be chosen.[6] 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:

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

where Υ has been defined as:

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

[6] 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 $p'(e_t)\Upsilon$ captures the effects on the Lagrangian of the government's problem caused by the increase in the unitary price of child care services due to the higher quality level.

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

$$MRS_{ec}^{j,t} = \frac{\partial V_t^j}{\partial e_t} / \frac{\partial V_t^j}{\partial B_t^j} = \left(H^2 - H^1\right) \eta_t'^j \frac{\partial \pi^j}{\partial e_t} / \frac{\partial V_t^j}{\partial B_t^j}. \tag{41}$$

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 allows to rewrite the f.o.c. with respect to e_t as:

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

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:

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

From (33) we can easily see that at an optimum $\Upsilon = -\mu_t \sum_{j=1}^{2} d_t^j f_t^j$. Therefore, dividing by μ_t all terms in the previous equation and rearranging terms, we get:

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

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:[7]

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

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

$$\sum_{j=1}^{2}f_t^jMRS_{ec}^{j,t}=p'\left(e_t\right)\sum_{j=1}^{2}d_t^jf_t^j+\frac{1}{\mu_t}\sum_{j=1}^{2}\left[\left(1-\varepsilon\right)\left(H^2-H^1\right)\eta_t^{\prime j}-\gamma_t\right]\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0}f_t^j+$$
$$\frac{\lambda_t}{\mu_t}\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(\widehat{MRS}_{ec}^{2,t}-MRS_{ec}^{1,t}\right)-\tau_t\sum_{j=1}^{2}\left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0}f_t^j. \tag{44}$$

Eq. (44) can be interpreted as a sort of modified Samuelson-type condition, even if it is not about 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

[7] 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.

(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 where: i) asymmetric information problems were absent (non self-selection constraints in the government's problem); ii) the government's objective function were welfaristic, which means that there were no laundering ($\varepsilon = 1$); iii) externalities were absent, in the sense that there were 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 allocation of time decisions. 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, in this way pushing 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.[8] In terms of 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 tightens 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 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.

4.1 Extension

A possible extension of the model that we have analyzed above 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 will 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 will also assume here that 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. Under the aforementioned assumptions, the design problem of the government is summarized by the following Lagrangian:

[8] 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.

$$
\begin{aligned}
\pounds \quad = \quad & \sum_{t=0}^{\infty}\rho^{t}\left[V_{t}^{1}+\varepsilon\eta\left(\pi^{1}\left(n_{t}^{1}\right)H^{2}+\left(1-\pi^{1}\left(n_{t}^{1}\right)\right)H^{1}\right)-\eta\left(\overline{\pi}^{1}\left(n_{t}^{1}\right)H^{2}+\left(1-\overline{\pi}^{1}\left(n_{t}^{1}\right)\right)H^{1}\right)\right]f_{t}^{1}+\\
& \sum_{t=0}^{\infty}\rho^{t}\left[V_{t}^{2}-\left(1-\varepsilon\right)\eta\left(\pi^{2}\left(n_{t}^{2}\right)H^{2}+\left(1-\pi^{2}\left(n_{t}^{2}\right)\right)H^{1}\right)\right]f_{t}^{2}+\\
& \sum_{t=0}^{\infty}\rho^{t}\mu_{t}\sum_{j=1}^{2}\left(Y_{t}^{j}-B_{t}^{j}+\tau_{t}d_{t}^{j}\right)f_{t}^{j}-\sum_{t=0}^{\infty}\rho^{t}\gamma_{t}\left[f_{t+1}^{2}-\sum_{j=1}^{2}f_{t}^{j}\pi^{j}\left(n_{t}^{j},e\right)\right]+\\
& \sum_{t=0}^{\infty}\rho^{t}\lambda_{t}\left(V_{t}^{2}-\widehat{V}_{t}^{2}\right).
\end{aligned}
$$

To save space, we will only consider how the new assumption that we have introduced affects 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 in this modified setting. 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 the following result:

$$
\begin{aligned}
METR_{t}^{1} \quad = \quad & \frac{\lambda_{t}}{\mu_{t}f_{t}^{1}}\frac{\partial\widehat{V}_{t}^{2}}{\partial B_{t}^{1}}\left(\frac{\frac{\partial\widehat{V}_{t}^{2}}{\partial Y_{t}^{1}}}{\frac{\partial\widehat{V}_{t}^{2}}{\partial B_{t}^{1}}}-\frac{\frac{\partial V_{t}^{1}}{\partial Y_{t}^{1}}}{\frac{\partial V_{t}^{1}}{\partial B_{t}^{1}}}\right)-\frac{\gamma_{t}}{\mu_{t}}\frac{\partial\pi^{1}}{\partial n_{t}^{1}}\left(\frac{dn_{t}^{1}}{dY_{t}^{1}}\right)_{dV_{t}^{1}=0}+\\
& \frac{1}{\mu_{t}}\left[\frac{\partial\overline{\pi}^{1}}{\partial n_{t}^{1}}\eta_{t}^{\prime 1}\left(\overline{EH}\right)-\varepsilon\frac{\partial\pi^{1}}{\partial n_{t}^{1}}\eta_{t}^{\prime 1}\left(EH\right)\right]\left(H^{2}-H^{1}\right)\left(\frac{dn_{t}^{1}}{dY_{t}^{1}}\right)_{dV_{t}^{1}=0}, \qquad (45)
\end{aligned}
$$

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 $\pi^1(\cdot)$.

The easiest way to understand the difference between (45) and (30) is to consider the no-laundering case where $\varepsilon = 1$. In that case eq. (30) would simplify to the first line of (45). Thus, the difference between (45) and (30) would be given by the presence in the former of the following additional term:

$$
\frac{1}{\mu_{t}}\left[\frac{\partial\overline{\pi}^{1}}{\partial n_{t}^{1}}\eta_{t}^{\prime 1}\left(\overline{EH}\right)-\frac{\partial\pi^{1}}{\partial n_{t}^{1}}\eta_{t}^{\prime 1}\left(EH\right)\right]\left(H^{2}-H^{1}\right)\left(\frac{dn_{t}^{1}}{dY_{t}^{1}}\right)_{dV_{t}^{1}=0}. \qquad (46)
$$

The term (46) 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 (46) 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 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 $\partial\overline{\pi}^1/\partial n_t^1 > \partial\pi^1/\partial n_t^1$. This effect pushes in the direction of making (46) negative (taking into account that $\left(dn_t^1/dY_t^1\right)_{dV_t^1=0} < 0$). This 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.

This is however not the end of the story since we have also to consider the relation between $\eta_t'^1\left(\overline{EH}\right)$ and $\eta_t'^1\left(EH\right)$. In this case, given our assumption that 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 seems reasonable to assume that $\overline{\pi}^1(n_t^1) > \pi^1(n_t^1)$ and therefore $\overline{EH} > EH$. Given that the η-function is increasing and concave, this implies $\eta_t'^1\left(\overline{EH}\right) < \eta_t'^1\left(EH\right)$, pushing in the direction of making (46) positive. This effect works in the opposite direction to the one that we have singled out above. The fact that low-skilled parents do not recognize the true shape of the function $\pi^1(n_t^1)$ lead them to underestimate the true marginal warm-glow effect of increasing n_t^1. This in turn represents an argument for weakening the low-skilled agents' incentives to provide labor by raising the marginal effective tax rate faced by them.

5 Concluding remarks

In the paper we characterize the optimal tax policy and the optimal level of quality of day care in a two-type OLG model with exogenous growth 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. Our paper has some similarities with Cremer and Pestieau (2006), who 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 affects the level of human capital of the respective offspring depends on the parents' skills in a non-monotonic way: children of low skilled individuals benefit from day-care as the quality of day-care centers is higher than the human capital of the parents; the opposite holds for kids of skilled parents. We show that this assumption could theoretically have important implications for the design of the optimal tax system.

A quantitative assessment of such implications represents a natural development of the analysis presented in the paper. Another interesting extension concerns the use of a two sector model to allow for a more detailed study of the production of day care. These issues are on our research agenda.

References

[1] Andreoni, J. (1989). Giving with impure altruism: application to charity and Ricardian equivalence. Journal of Political Economy, 97, 1447-1458.

[2] Atkinson, A. B. and J. E. Stiglitz (1976). The design of tax structure: direct versus indirect taxation. Journal of Public Economics, 6, 55-75.

[3] Banerjee, A. V. and A. F. Newman (1991). Risk-bearing and the theory of income distribution. Review of Economic Studies, 58, 211-235.

[4] Bernal, R. (2008). The effect of maternal employment and child care on children's cognitive development. International Economic Review, 49, 1173-1209.

[5] Bernal, R. and M. Keane (2007). Child care choices and children's cognitive achievement: the case of single mothers. mimeo.

[6] Bernal, R. and M. Keane (2009). Quasi-structural estimation of a model of child-care choices and child cognitive ability production. Journal of Econometrics. (forthcoming).

[7] Carneiro, P., Meghir, C. and M. Parey (2007). Maternal education, home environments and the development of children and adolescents. CEPR Discussion Paper 6505.

[8] Carneiro, P. and J. Heckman (2003). Human capital policy. NBER wp 9495.

[9] Casarico A. and A. Sommacal (2008). Labour income taxation, Human capital and growth: the role of child care, CEPR discussion paper n. 7039.

[10] Cremer, H. and P. Pestieau (2006). Intergenerational transfer of human capital and optimal education policy. Journal of Public Economic Theory, 8, 529-545.

[11] Cunha, F. and J. Heckman (2007). The technology of skill formation. American Economic Review, 97, 31-47.

[12] 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.

[13] Edwards, J. Keen, M. and M. Tuomala (1994). Income tax, commodity taxes and public good provision: a brief guide. FinanzArchiv, 51, 472-487.

[14] Guryan, J., Hurst, E. and M. Schettini Kearney (2008). Parental education and parental time with children. Journal of Economic Perspectives, 22, 23-46.

[15] Kimmel, J. and R. Connelly (2007). Mothers' time choices: caregiving, leisure, home production, and paid work. Journal of Human Resources, 42, 643-661.

[16] 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.

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