CESifo Working Papers

www.cesifo.org/wp

Intergenerational Transmission of Skills during Childhood and Optimal Public Policy

Alessandra Casarico
Luca Micheletto
Alessandro Sommacal

CESIFO WORKING PAPER NO. 3343
CATEGORY 1: PUBLIC FINANCE
FEBRUARY 2011

PRESENTED AT CESIFO AREA CONFERENCE ON EMPLOYMENT AND SOCIAL PROTECTION, MAY 2010

CESifo *Working Paper No. 3343*

Intergenerational Transmission of Skills during Childhood and Optimal Public Policy

Abstract

The paper characterizes the optimal tax policy and the optimal quality of day care services in a OLG model with warm-glow altruism where parental choices over child care arrangements affect the probability that the child becomes a high-skilled adult in a type-specific way. With respect to previous contributions, optimal tax formulas include type-specific Pigouvian terms which correct for the intergenerational externality in human capital accumulation. Our numerical simulations suggest that a public policy that disregards the effects of parental time on children's human capital entails a welfare loss that ranges from 0:2% to 5:7% of aggregate consumption.

JEL-Code: H210, H230, J130, J220, J240.

Keywords: optimal taxation, day care quality, intergenerational transmission of skills, early childhood environment, warm-glow.

Alessandra Casarico
Bocconi University
Milan / Italy
alessandra.casarico@unibocconi.it

Luca Micheletto
Faculty of Law
University of Milan / Italy
luca.micheletto@unibocconi.it

Alessandro Sommacal
Faculty of Economics
University of Verona / Italy
alessandro.sommacal@unibocconi

The paper was presented at the 2009 SIEP conference in Pavia, at the University of Louvain, at the University of St.Gallen, at the University of Wien, at the 2010 CESifo Area Conference on Employment and Social protection and at the 2010 IIPF Conference in Uppsala. We thank all the conference and seminar participants for their useful comments and suggestions. All the remaining errors are ours.

1 Introduction

The paper aims at characterizing the optimal tax policy and the optimal quality of day care services in an OLG model where parental choices over child care arrangements (that is, parental time devoted to children and time spent in day care centers) affect in a type-specific way the probability that a child grows up as a high-market-ability adult.

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 Cunha 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 et al. 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. 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 (2009, 2010) 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 alternative sources of child care and levels of maternal education: for instance, formal care (i.e. center-based care and preschool) may have positive effects on children of poorly educated mothers.[1] 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. 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' 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.[2] We assume that agents are heterogeneous in two dimensions: market ability and ability to raise children. Both abilities can

[1] Havnes and Mogstad (2010) show that the focus on mean impact of day care services on child development can also mask important differences along the earnings distribution. The effects are positive and sizable below the median of the earnings distribution.

[2] 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 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).

be either high or low. 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. 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. Moreover, we assume that the human capital transmission is type-specific, in the sense that the effect of child care arrangements on the probability for a child to become a high-market-ability adult depends on the parent's type. We consider two different scenarios: first, one where the government can use linear taxation on labor income and a linear tax/subsidy on the purchase of day care services. 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 model. A high (low) market ability type will in this case also have a high (low) ability to raise children. Under this circumstance we will focus on the case 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 have their probability of being high skilled tomorrow reduced, by the substitution of parental time with day care.[3] In both frameworks, we discuss the rules dictating the optimal choice of day care quality enforced by the government.

With respect to previous contributions, we find that optimal tax formulas incorporate type-specific Pigouvian terms which correct for the intergenerational externality in human capital accumulation. 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 depends on the parent's type. When only linear instruments are available to the government, the tax rates 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 a nonlinear tax on labor income is at the government's disposal, the government can let agents face type-specific marginal income tax rates. This is helpful since it allows the government to better tailor the distortion imposed on each agent to the externality that he generates.

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

[3] This assumption is further discussed in Section 4.

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, this is determined by equating the total private marginal benefits of a quality increase to the marginal cost adjusted for the presence of three additional terms: the first one captures the impact on the government's budget constraint through the change in demand of consumption and day care services. The second one reflects the intergenerational externality in human capital accumulation. The third one, which is present only in the mixed tax system, comes from the self-selection constraint.

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 done, for instance, in Cremer and Pestieau (2006). 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.

Admittedly, the inclusion of day care services into a second-best optimal taxation framework is not a novelty of our analysis (see e.g. Cremer and Gahvari 1997, Blomquist et al. 2010 and Blomquist and Micheletto 2009). However, when the optimal taxation literature discusses day care services, it typically does so by treating them just as one prominent example 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. 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. Child care enters the human capital production function in Casarico and Sommacal (2011) which studies the impact on growth of changes in labor income taxation. However, their analysis is not concerned with the design of optimal public policies.

The paper is organized as follows. Section 2 presents the basic ingredients of the model and provides a description of the behavior of agents, the productive technology, the evolution over time of the skill distribution in the population and the government's objective function. Section 3 analyzes the solution to the government's problem under a linear tax system, whereas a mixed tax system is considered in Section 4. Section 5 discusses some possible extensions of our model and Section 6 provides a numerical example aimed at illustrating the magnitude of the welfare loss from designing the public policy without taking into account the effects of parental time on children's future human capital. Finally, Section 7 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 below, the distribution of the ability to raise children 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, each agent has a child and, given his level of skills, he decides how to allocate time between labor, time devoted to children and leisure. Denoting time indices by a subscript, a parent of market ability type j and ability to raise children type k (hereafter labelled simply as parent of ability type jk) maximizes the following expected utility function:

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

with $u''(\cdot) < 0 < u'(\cdot)$. We denote by c_t^{jk}, z_t^{jk} and n_t^{jk} respectively consumption, leisure time without kids and leisure time spent with kids by a jk-agent. The term $\eta(H^j)$ reflects the warm-glow altruism of parents (Andreoni 1989) towards the level of market ability of their children H^j, $j = 1, 2$, with $H^2 > H^1$ and $\eta'(\cdot) > 0$. As regards the ability to raise children, we assume that it can only take two possible values, low and high, respectively denoted by $k = 1$ and $k = 2$.[4] 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. $\pi^{jk}(n_t^{jk})$ is the probability of having a high-human capital child and it stands for $\pi^k(n_t^{jk}, H^j, e_t)$: the probability of 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 day care services e_t, which individuals take as given and which we treat as a choice variable for the government.

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

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

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

[4] 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.

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 the sake of exposition that $\overline{a} = 1$.

We assume that for any given n_t the following condition holds: $\pi_n^{11}(n_t) < \pi_n^{jk}(n_t) < \pi_n^{22}(n_t)$, (or, equivalently given the time constraint (3), $\pi_d^{22}(d_t) < \pi_d^{jk}(d_t) < \pi_d^{11}(d_t)$), with $j \neq k$. These inequalities imply that an increase in the time spent with children by an agent who has low ability to raise children and low market ability is less beneficial for the child's probability of becoming a high-market-ability adult 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.

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), determines f_t^{jk} for any j, k and t.

2.4 Government

As to the government, the objective function is:

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

where ρ is the social planner discount factor. The parameter $\varepsilon \in [0,1]$ allows the government to launder out the warm-glow component.[5] When $\varepsilon = 1$, the government fully takes into account the individual preferences; when $\varepsilon = 0$, the government fully launders out the warm-glow component.

As to the government's budget constraint, 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 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. Thus, we can write the agents' budget constraint as:

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

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

The budget constraint for the government, which we assume to be balanced year by year without recurring to debt, can be written as:

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

3.1 Solution to the consumer optimization problem and indirect utility function

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

$$\begin{aligned} u'_{c_t^{jk}} &= \varkappa_t^{jk}(1+\tau_t^c), & (10)\\ -u'_{d_t^{jk}} + \frac{\partial \pi^{jk}}{\partial d_t^{jk}}(\eta(H^2)-\eta(H^1)) &= \varkappa_t^{jk}(p(e_t)+\tau_t^d - wH^j), & (11)\\ u'_{z_t^{jk}} &= \varkappa_t^{jk} wH^j, & (12) \end{aligned}$$

[5] How the warm-glow component should 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).

where $u'_{y_t} = \frac{\partial u}{\partial y_t}$ and where $\varkappa_t^{jk}$ is the Lagrangian multiplier which 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 $V_t^{jk} = u\left(c_t^{jk*}, z_t^{jk*}, n_t^{jk*}\right) + \pi^{jk}(n_t^{jk*})\eta(H^2) + (1 - \pi^{jk}(n_t^{jk*}))\eta(H^1)$, 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).

3.2 The government's problem and the optimal policy

The government maximizes:

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

with respect to τ_t^x, G_t and e_t, where μ_t and υ_t are the multipliers of the constraints (9) and (5).

We begin our analysis of the optimal policy by characterizing the optimal tax structure. For this purpose it is convenient to introduce the concept of net marginal social evaluation of an agent's income. Denoting by b_t^{jk} the net marginal social evaluation of agent jk's income, we have:

$$b_t^{jk} = \frac{\varkappa_t^{jk}}{\mu_t} + \tau_t^c \frac{\partial c_t^{jk}}{\partial G_t} + \tau_t^d \frac{\partial d_t^{jk}}{\partial G_t} - \frac{\partial \pi^{jk}}{\partial d_t^{jk}} \frac{\partial d_t^{jk}}{\partial G_t} \frac{1}{\mu_t}\left[(1-\varepsilon)(\eta(H^2) - \eta(H^1)) - \upsilon_t\right]. \qquad (14)$$

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 and third terms indicate the impact on the tax revenues associated with the change in the demand for the two goods. The fourth term shows the impact that a change in the lump-sum transfer has on the demand for day 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 υ_t. When $\varepsilon \neq 1$, the social evaluation will also depend on the degree of laundering out.

We are now ready to characterize the optimal tax structure.

Proposition 1 *Denoting Hicksian demands by a "tilde", and the expectation and covariance operators respectively by $E(\cdot)$ and $Cov(\cdot)$, the optimal linear tax system is characterized by the following set of conditions:*

$$E(b_t) = 1, \qquad (15)$$

$$\frac{\tau_t^c \sum_{j=1}^2 \sum_{k=1}^2 f_t^{jk} \partial \widetilde{x}_t^{jk} / \partial \tau_t^c + \tau_t^d \sum_{j=1}^2 \sum_{k=1}^2 f_t^{jk} \partial \widetilde{x}_t^{jk} / \partial \tau_t^d}{x_t}$$

$$= -\left[1 - Cov(b_t^{jk}, \frac{x_t^{jk}}{x_t})\right] + \qquad (16)$$

$$\frac{1}{\mu_t} \sum_{j=1}^2 \sum_{k=1}^2 \frac{\partial \pi^{jk}}{\partial d_t^{jk}} \frac{\partial \widetilde{d}_t^{jk} / \partial \tau_t^x}{x_t} f_t^{jk} \left[(1-\varepsilon)(\eta(H^2) - \eta(H^1)) - \upsilon_t \right],$$

for $x = c, d$.

Proof. See Appendix A. ■

Condition (15) defines the optimal level of the uniform lump-sum transfer and has a standard interpretation. It requires that at an optimum the lump-sum transfer should be adjusted such that b, the government's net marginal valuation of a transfer of 1\$ (measured in terms of government's revenue), should on average be equal to its marginal cost.

Condition (16) characterizes the optimal tax rates. On the left hand side we have the proportional change in the aggregate compensated demand for good x due to indirect taxes. This 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 child care arrangements have on human capital accumulation. In this new term we can identify two components: the first one, that is $\frac{1}{\mu_t} \sum_{j=1}^2 \sum_{k=1}^2 \frac{\partial \pi^{jk}}{\partial d_t^{jk}} \frac{\partial \widetilde{d}_t^{jk} / \partial \tau_t^x}{x_t} f_t^{jk} \left[(1-\varepsilon)(\eta(H^2) - \eta(H^1)) \right]$, 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 $-\frac{1}{\mu_t} \sum_{j=1}^2 \sum_{k=1}^2 \frac{\partial \pi^{jk}}{\partial d_t^{jk}} \frac{\partial \widetilde{d}_t^{jk} / \partial \tau_t^x}{x_t} f_t^{jk} \upsilon_t$, identifies the externality related to the assumption of impure 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 tax instruments available to the government are linear, which implies that the tax rates applied to the two goods need to be the same across individuals, 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. If we assume that $\partial \pi^{11} / \partial d_t^{11} > 0$ while $\partial \pi^{22} / \partial d_t^{22} < 0$,[6] the correction ideally imposed on these two types is of opposite sign. Thus, if we consider for example the case of $\varepsilon = 1$, we will have that the adjustment ideally required to correct the time allocation of a parent of type 11 (resp.: 22) would call for not discouraging (resp.: not encouraging) 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 derivatives, namely $\partial \pi^{jk} / \partial d_t^{jk}$ when

[6] This appears a quite natural assumption if one agrees that the quality of day care services cannot be lower than that provided at home by parents of type 11 and it cannot be higher than that provided at home by parents of type 22.

$j \neq k$. It seems reasonable to believe that the sign of the derivative will in these cases crucially depend on the value of time spent in day care centers 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.[7] For this reason, we avoid making specific assumptions on the sign of $\partial\pi^{jk}/\partial d_t^{jk}$ for $j \neq k$. Depending on the sign taken by each of these derivatives, 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. Notice that the weight of each of these terms will depend on how strong the effect of the tax rate is on the compensated demand for day care services of the different agents' types.

We now turn our attention to how the quality of day care services should be chosen. Proposition 2 provides the main result.

Proposition 2 *Under a linear tax system the optimal quality of day care services abides by the following rule:*

$$\begin{aligned}\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left(\frac{\partial V_t^{jk}}{\partial e_t} - p'(e_t)\,\mu_t d_t^{jk}\right) &= -\mu_t\left(\tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial c_t^{jk}}{\partial e_t} + \tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial d_t^{jk}}{\partial e_t}\right) \\ &\quad + \sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left[(1-\varepsilon)\left(\eta(H^2)-\eta(H^1)\right)-\upsilon_t\right]\left(\frac{\partial\pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial e_t} + \frac{\partial\pi^{jk}}{\partial e_t}\right).\end{aligned} \tag{17}$$

Proof. See Appendix A. ■

The left-hand side evaluates the effect on the agents' utilities of a marginal increase in the quality of day care services. For each type of agent, it is given by the difference between the effect on utility (keeping constant the consumer price $p(e_t) + \tau_t^d$) and the effect on the agents' monetary outlays, keeping constant the demand for day care services. The first term on the right hand side measures the impact on tax revenues of a higher quality of day care. The second term takes into account that a 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 impure altruism.

[7] 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.

Having completed the analysis of the public policy under linear tax instruments, we devote the next section to investigate the government's problem under the assumption that earned income can be taxed on a nonlinear scale.

4 Mixed tax system

In this section we let the government optimize the design of a tax structure where earned income is subject to a nonlinear income tax whereas the purchase of goods/services is subject to a set of differentiated linear commodity taxes.

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 to day care services.

With the government being able to observe earned income at an individual level but unable to observe an individual's labor supply and his wage rate, the tax design problem 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 the well-known complexities related to multidimensional screening problems, in this section we assume perfect correlation between the two types of abilities, i.e. market ability and ability to raise children. Under this assumption, we have 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.

Defining by $B_t^j \equiv Y_t^j - T_t\left(Y_t^j\right)$ the net income of agent j, 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 public budget constraint and a set of self-selection constraints. Following the bulk of the literature, we 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.

Denoting by V_t^j the maximum utility that can be attained by a type j agent who chooses the (Y, B)-bundle intended for him by the government, and by $\widehat{V}_t^2$ the maximum utility that can be attained by a high ability mimicker, we have (remembering that $d_t^j = 1 - n_t^j$):

$$\begin{aligned}
V_t^j &= V\left(Y_t^j, B_t^j; w_t H^j\right) \\
&= \max_{n_t^j}\left\{u\left(B_t^j - (p(e_t)+\tau_t)\left(1-n_t^j\right), 1 - \frac{Y_t^j}{w_t H^j} - n_t^j, n_t^j\right) + \pi^j(n_t^j)\eta(H^2) + (1-\pi^j(n_t^j))\eta(H^1)\right\}; \\
\widehat{V}_t^2 &= V\left(Y_t^1, B_t^1; w_t H^2\right) \\
&= \max_{\widehat{n}_t^2}\left\{u\left(B_t^1 - (p(e_t)+\tau_t)\left(1-\widehat{n}_t^2\right), 1 - \frac{Y_t^1}{w_t H^2} - \widehat{n}_t^2, \widehat{n}_t^2\right) + \pi^2(\widehat{n}_t^2)\eta(H^2) + (1-\pi^2(\widehat{n}_t^2))\eta(H^1)\right\}.
\end{aligned}$$

The design problem can be summarized by the following Lagrangian:

$$\begin{aligned}
\pounds \;=\; & \sum_{t=0}^{\infty}\rho^t \sum_{j=1}^{2} f_t^j \left\{V_t^j - (1-\varepsilon)\left[\pi^j\left(n_t^j\right)\eta\left(H^2\right) + \left(1-\pi^j\left(n_t^j\right)\right)\eta\left(H^1\right)\right]\right\} + \\
& \sum_{t=0}^{\infty}\rho^t \mu_t \sum_{j=1}^{2} f_t^j\left(Y_t^j - B_t^j + \tau_t d_t^j\right) - \sum_{t=0}^{\infty}\rho^t \upsilon_t\left[f_{t+1}^2 - \sum_{j=1}^{2} f_t^j \pi^j\left(n_t^j\right)\right] + \\
& \sum_{t=0}^{\infty}\rho^t \lambda_t\left(V_t^2 - \widehat{V}_t^2\right). \qquad (18)
\end{aligned}$$

As we have done in the previous section, we begin by characterizing the optimal tax structure and then move to the analysis of the optimal quality of day care services.

4.1 The optimal tax structure

As a measure of the distortions imposed by an optimal tax structure on the agents' labor supply we will focus here on the concept of marginal effective tax rate ($METR$). This is defined as the variation in total (income and commodity) taxes paid by an agent if he were to earn an additional unit of gross income. Formally, the marginal effective tax rate faced by agents of type i ($i = 1, 2$) is defined as:

$$METR_t^i \equiv T'\left(Y_t^i\right) + \left[\frac{\partial d_t^i}{\partial Y_t^i} + \left(1 - T'\left(Y_t^i\right)\right)\frac{\partial d_t^i}{\partial B_t^i}\right]\tau_t, \qquad (19)$$

where $T'\left(Y_t^i\right)$ denotes the marginal income tax rate faced by agents of type i.

Since agents choose labor supply maximizing $V\left(Y_t^j, B_t^j; w_t H^j\right)$ subject to the link between pre-tax earnings and post-tax earnings available for goods expenditures implied by the direct tax schedule, an optimizing behavior implies $\frac{\partial V_t^i}{\partial B_t^i}\left(1 - T'\left(Y_t^i\right)\right) + \frac{\partial V_t^i}{\partial Y_t^i} = 0$. This allows defining the (implicit) marginal income tax rate faced by an agent as:

$$T'\left(Y_t^i\right) = 1 + \left(\frac{\partial V_t^i}{\partial Y_t^i} / \frac{\partial V_t^i}{\partial B_t^i}\right). \qquad (20)$$

Thus, we can rewrite (19) as:

$$METR_t^i \equiv T'\left(Y_t^i\right) + \left(\frac{dd_t^i}{dY_t^i}\right)_{dV_t^i=0} \tau_t, \tag{21}$$

where we have defined $\left(\frac{dd_t^i}{dY_t^i}\right)_{dV_t^i=0} \equiv \frac{\partial d_t^i}{\partial Y_t^i} - \left(\frac{\partial V_t^i}{\partial Y_t^i} / \frac{\partial V_t^i}{\partial B_t^i}\right) \frac{\partial d_t^i}{\partial B_t^i}$. Notice that, since $d_t^i = 1 - n_t^i = l_t^i + z_t^i = \frac{Y_t^i}{wH^i} + z_t^i$, we also have $\left(\frac{dd_t^i}{dY_t^i}\right)_{dV_t^i=0} = \frac{1}{wH^i} + \left(\frac{dz_t^i}{dY_t^i}\right)_{dV_t^i=0}$ and $\left(\frac{dd_t^i}{dY_t^i}\right)_{dV_t^i=0} = -\left(\frac{dn_t^i}{dY_t^i}\right)_{dV_t^i=0}$.

We are now ready to provide the following result.

Proposition 3 *Under a mixed tax system the optimal marginal effective tax rates are given by:*

$$METR_t^2 = \frac{1}{\mu_t}\left[(1-\varepsilon)\left(\eta\left(H^2\right) - \eta\left(H^1\right)\right) - \upsilon_t\right] \frac{\partial \pi^2}{\partial n_t^2} \left(\frac{dn_t^2}{dY_t^2}\right)_{dV_t^2=0}, \tag{22}$$

$$METR_t^1 = \frac{\lambda_t}{\mu_t f_t^1} \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \left(\frac{\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} - \frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\right) + \frac{1}{\mu_t}\left[(1-\varepsilon)\left(\eta\left(H^2\right) - \eta\left(H^1\right)\right) - \upsilon_t\right] \frac{\partial \pi^1}{\partial n_t^1} \left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0}. \tag{23}$$

Proof. See Appendix B. ■

Starting with (22), the sign of $\partial \pi^j / \partial n_t^j$ could in general be either positive or negative. However, since (22) refers to agents who possess both high market ability and high ability to raise children, we regard the case when $\partial \pi^2 / \partial n_t^2 > 0$ as more relevant. If this is the case, and 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 to z and n goes down), the sign of (22) 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 given 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 that they become high-skilled 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 (23), which provides an expression for the METR faced by low-skilled agents at an optimum. The second term on the right hand side of (23) has the same structure as the term appearing on the right hand side of (22) and it can be interpreted in a similar way. However, since (23) refers to agents who possess both low market ability and low ability to raise children, in this case we regard the case when $\partial\pi^1/\partial n_t^1 < 0$ as more relevant.[8] Thus, the sign of the second term on the right hand side of (23) 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 (23) 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 and for time with the children. 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 (23) 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 (23) 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. This distortion is related to the assumption that the government cannot observe "who is who" and it is therefore constrained to design the income tax schedule subject to a self-selection constraint. The sign of this self-selection term coincides with the sign of the expression within brackets. In standard models of nonlinear redistributive income taxation,[9] a so-called agent-monotonicity assumption is usually invoked for the purpose of signing the distortion produced by the self-selection term. 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. Under this assumption the sign of the first term on the right side of (23) is positive, therefore calling for a downward distortion on the labor supply of low-skilled agents.[10]

[8] It is important to 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. What we are assuming here is that, at the margin and in the neighborhood of a social optimum, a substitution of time spent with children for time spent in day care centers lowers the probability that children of low skilled parents become high-skilled adults.

[9] See, for example, Stiglitz (1982) or Edwards et al. (1994).

[10] Notice however that the conditions required to satisfy the agent-monotonicity assumption are stronger in our setting than in standard optimal taxation models. In the latter, normality of consumption is a sufficient

Having characterized the optimal distortions imposed by taxation on the labor supply of the different types of agents, we can look at the optimal tax/subsidy on day care expenditures. Denoting Hicksian demands by a "tilde", the next Proposition provides the main result.

Proposition 4 *Under a mixed tax system the optimal tax rate on day care expenditures is given by:*

$$\tau_t = \frac{\lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(d_t^1 - \widehat{d}_t^2\right) + \sum_{j=1}^{2}\left[(1-\varepsilon)\left(\eta\left(H^2\right) - \eta\left(H^1\right)\right) - \upsilon_t\right] \frac{\partial \pi^j}{\partial d_t^j} \frac{\partial \widetilde{d}_t^j}{\partial \tau_t} f_t^j}{\mu_t \sum_{j=1}^{2} \frac{\partial \widetilde{d}_t^j}{\partial \tau_t} f_t^j}. \tag{24}$$

Proof. See Appendix B. ■

The denominator of the expression on the right hand side of (24) 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 (24). The first term at the numerator depends on the difference between the amount of day care services used by a true low-skilled and by a high-skilled mimicker. As we have already 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 hand side of (24) 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 day 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 (24) 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

$$\tau_t = \frac{\lambda_t}{\mu_t} \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \frac{d_t^1 - \widehat{d}_t^2}{\sum_{j=1}^{2} \frac{\partial \widetilde{d}_t^j}{\partial \tau_t} f_t^j} - \frac{1}{\mu_t} \sum_{j=1}^{2} \left[(1-\varepsilon)\left(\eta\left(H^2\right) - \eta\left(H^1\right)\right) - \upsilon_t\right] \frac{\partial \pi^j}{\partial n_t^j} \zeta_t^j. \tag{25}$$

Written in this form, the second term on the right hand side of (25) is reminiscent of a similar term appearing in (22) and (23). The main difference is that in (25) we take a sum over $j = 1, 2$ whereas in both (22) and (23) 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

condition for agent-monotonicity. In our setting, this is not 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). See Appendix C for further details.

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. But for the purpose of letting agents internalize the social effect of their time spent with their 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 that, since $\zeta_t^j > 0$ for all j but the sign of $\partial\pi^j/\partial n_t^j$ is assumed to be type-specific, the direction of the required adjustment in τ_t will be the opposite for high- and low-skilled agents. Thus, at least for the no-laundering scenario or for small degree of laundering (values of ε which are close to one), 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 for day care services by agents of type j. 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.[11]

Having characterized the structure of the optimal mixed tax system, we can notice that the so called "principle of targeting" fails to hold in this set up. This principle states that a distortion is best addressed by the instrument that acts directly on the relevant margin.[12] In our setting the time spent by parents with their kids 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

[11] Notice that a system often used in practice is to let the tax/subsidy rate on day care expenditures be a function of the income earned by parents. The reason why we do not consider this possibility here is that, with respect to the analysis presented above, letting τ_t be a function of Y_t would produce only minor differences. In particular, the expressions for the optimal marginal *effective* tax rates faced by the various types of agents would still be given by (22) and (23). Moreover, the marginal *income* tax rates $T'\left(Y_t^j\right)$ faced by the various types of agents would not be uniquely defined given that it would always be possible for the government to compensate a decrease in $T'\left(Y_t^j\right)$ with an increase in $\tau'\left(Y_t^j\right)$.

[12] See Dixit (1985).

depending on ε and υ_t do not vanish from the expressions characterizing the marginal tax rates for high- and low-skilled agents even after substituting in (45) and (49) the expression for τ_t provided by (25).

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 the one hand the government would like to distort the demand for day care services coming from high- and low-skilled parents in different ways. 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 this allows the government to have different agents facing different marginal tax rates. Under this respect, the nonlinear income tax is designed also to serve the role of an (imperfect) substitute for the impossibility to tax nonlinearly the agents' purchases of day care services.[13]

We can now turn our attention to the quality of day care services.

4.2 The optimal quality of day care services

Defining by $MRS_{ec}^{j,t}$ the marginal rate of substitution between the quality of day care services and private consumption (keeping fixed the consumer price of day care services) for an agent of type j at time t, we have:

$$MRS_{ec}^{j,t} \equiv \frac{\partial V_t^j}{\partial e_t} / \frac{\partial V_t^j}{\partial B_t^j} = \left[\eta\left(H^2\right) - \eta\left(H^1\right)\right] \frac{\partial \pi^j}{\partial e_t} / \frac{\partial V_t^j}{\partial B_t^j}. \tag{26}$$

With the help of (26) Proposition 5 provides a characterization of the optimal quality of day care services under a mixed tax system.

Proposition 5 *Under a mixed tax system the optimal quality of day care services abides by*

[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. Notice that such a system is informationally different from a system where the tax/subsidy rate on day care services is income-dependent. A fully nonlinear tax system requires the public authority to be able to observe the purchases of day care services at an individual level; a system where the tax/subsidy rate on day care services is income-dependent only requires the government to be able to observe earned income at the individual level. In terms of results, the main difference between a fully nonlinear tax system and the mixed tax system that we have analyzed in the text would be that, under a fully nonlinear 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.

the following rule:

$$\sum_{j=1}^{2} f_t^j MRS_{ec}^{j,t} = p'(e_t)\sum_{j=1}^{2} d_t^j f_t^j + \frac{1}{\mu_t}\sum_{j=1}^{2}\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0} f_t^j +$$
$$\frac{\lambda_t}{\mu_t}\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(\widehat{MRS}_{ec}^{2,t} - MRS_{ec}^{1,t}\right) - \tau_t \sum_{j=1}^{2}\left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0} f_t^j, \tag{27}$$

where $\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0}$ *and* $\left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0}$ *are defined as:*

$$\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0} \equiv \frac{\partial \pi^j}{\partial e_t} + \left(\frac{\partial n_t^j}{\partial e_t} - \frac{\partial n_t^j}{\partial B_t^j}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}\right)\frac{\partial \pi^j}{\partial n_t^j}, \tag{28}$$

$$\left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0} \equiv \frac{\partial d_t^j}{\partial e_t} - \frac{\partial d_t^j}{\partial B_t^j}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}. \tag{29}$$

Proof. See Appendix B. ■

Equation (27) can be interpreted as a 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 equation (27) measures the sum of the agents' marginal willingness to pay for an increased level of quality of day care services. The first term on the right hand side of (27) represents the additional resource cost of raising the quality of day care services, keeping the consumption of services by agents fixed. It is the only term that would remain in a setting 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 day 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 (27) 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 (27) does not challenge this interpretation because, as evident from (25), 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 (27) 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 day 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 day 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 ($\varepsilon \rightarrow 1$), we can conclude that the sign of the second term on the right side of (27) is negative, which pushes for an increase in the second-best efficient level of day care quality.

The third term on the right side of (27) is a self-selection term that depends on the difference between a mimicker's marginal willingness to pay for increased day 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, (26) tends to imply that the marginal willingness to pay for increased quality is larger for a mimicker than for a true low-skilled.[14] 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 (27) 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 day care quality is implemented. Assuming that agents' consumption of day care services goes up when the quality of services increases, the last term on the right hand side of (27) raises (resp.: lowers) the net marginal cost of quality whenever the purchase 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 our baseline model, focusing on the mixed tax system case.

[14] It is however clear from (26) 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.

5.1 Faulty beliefs on π

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-ability adults. For illustrative purposes, we consider here the case where only low-ability 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 that the child will be a high-ability type. Moreover, we assume that low-ability agents tend to underestimate the negative effect that n_t^1 has on the expected human capital of their children. Under the aforementioned assumptions, the design problem of the government is summarized by the following Lagrangian:

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

To save space, we only consider the effects on the results about the optimal marginal effective tax rates faced by high- and low-ability agents.

With respect to the former, it is quite easy to realize that the result given by (22) is still valid. Intuitively, since we have not changed any assumption pertaining to the behavior of the high-ability agents, the structure of the optimal distortion imposed on them should remain unaffected. With respect to the low-ability agents, instead, the result provided by (23) 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 (23) above, it would only be a matter of tedious calculations to end up with:

$$\begin{aligned}
METR_t^1 \;=\; & \frac{\lambda_t}{\mu_t f_t^1}\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} - \frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\right) + \\
& + \frac{1}{\mu_t}\left[(1-\varepsilon)\left(\eta\left(H^2\right) - \eta\left(H^1\right)\right) - \upsilon_t\right]\frac{\partial \pi^1}{\partial n_t^1}\left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0} \\
& \underbrace{+\frac{\eta\left(H^2\right) - \eta\left(H^1\right)}{\mu_t}\left(\frac{\partial \overline{\pi}^1}{\partial n_t^1} - \frac{\partial \pi^1}{\partial n_t^1}\right)\left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0}}_{\Gamma}.
\end{aligned} \tag{30}$$

Comparing (30) and (23) 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-ability 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-ability 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-ability agents should be larger (resp.: smaller) when they misperceive the true shape of the function $\pi^1(\cdot)$.

Given our assumption that the low-ability 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 $\partial\overline{\pi}^1/\partial n_t^1 > \partial\pi^1/\partial n_t^1$. Thus, the sign of the bracketed difference in Γ takes a positive sign, reflecting the circumstance that low-ability 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-ability parents underestimate the negative effect of n_t^1 tends to make them spend too much time with their children, at least too much when compared with the time that would be spent by a utility-maximizing low-ability parent who correctly perceived the shape of $\pi^1(n_t^1)$. In this case, a lower marginal effective tax rate represents an instrument to distort the low-ability 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.

5.2 Peer effects

Another possible extension of the model analyzed in Section 4 is to introduce externalities across peers. 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 $\pi^j(n_t^j, H^j, e_t, n_t^{k\neq j})$, rather than by $\pi^j(n_t^j, H^j, e_t)$ as assumed so far. The externality we have in mind can be described as follows: children of low-ability agents benefit from the interaction in day care centers with children of high-ability agents, meaning that a prolonged interaction with children of high-ability agents increases their probability of becoming high-human-capital adults. For the sake of argument, we consider that the reverse is true for children of high-ability agents; namely, a prolonged interaction with children of low-ability agents in day care centers decreases their probability of becoming high-human-capital adults.

We can rewrite the government's problem as:

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

where we have defined $E_t^j = n_t^{k \neq j}$ the externality faced by agents of type j and where we have added two externality constraints, with associated multipliers θ_t^1 and θ_t^2, in the Lagrangian of the government.

If the children of low-ability agents benefit from the interaction in day care centers with children of high-ability agents, and if the reverse is true for children of high-ability agents, the sign of the multiplier θ_t^1 is positive, since n_t^2 generates a negative externality on low-ability agents, whereas the sign of the multiplier θ_t^2 is negative, since n_t^1 generates a positive externality on high-ability agents.

What would this imply for the conclusions of our model? Also here, to save space, we 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-ability agents would be respectively given by:

$$
METR_t^2 = \frac{1}{\mu_t} \left\{ \left[(1-\varepsilon) \left(\eta \left(H^2 \right) - \eta \left(H^1 \right) \right) - \upsilon_t \right] \frac{\partial \pi^2}{\partial n_t^2} + \frac{\theta_t^1}{f_t^2} \right\} \left(\frac{d n_t^2}{d Y_t^2} \right)_{dV_t^2 = 0}; \tag{31}
$$

$$
\begin{aligned}
METR_t^1 \; = \; & \frac{\lambda_t}{\mu_t f_t^1} \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \left(\frac{\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} - \frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}} \right) + \\
& \frac{1}{\mu_t} \left\{ \left[(1-\varepsilon) \left(\eta \left(H^2 \right) - \eta \left(H^1 \right) \right) - \upsilon_t \right] \frac{\partial \pi^1}{\partial n_t^1} + \frac{\theta_t^2}{f_t^1} \right\} \left(\frac{d n_t^1}{d Y_t^1} \right)_{dV_t^1 = 0}.
\end{aligned} \tag{32}
$$

From a formal point of view, the only difference with respect to the corresponding formulas of Section 4 (eqs. (22) and (23)) is the presence in (31) and (32) 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 (31) and (32), opposite to the sign of the υ-term, pushing in the direction of lowering the marginal effective tax rate faced by the high-ability agents and raising the marginal effective tax rate faced by the low-ability agents. The first effect is explained by noticing that a reduction in the marginal effective tax rate faced by the high-ability agents is a way to induce them to internalize the negative externality that, by spending more time with their own kids, they impose on low-ability agents: a lower marginal effective tax rate on high-ability 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 (32) calls for an increase in the marginal effective tax rate faced by low-ability agents.

6 Numerical example

In this section we provide a numerical example to illustrate the working of the model without considering the extensions developed in Section 5. Notice that we do not perform a detailed calibration on a specific country but we rather elaborate a numerical analysis to understand the relevance of the mechanisms studied in the theoretical part: how does the presence of a link between day care arrangements and the distribution of skills affect the optimal policies chosen by the government? How strong should such a link be, to play a sizable role in the design of the optimal policies?

To address these issues, we compare two models. In what we call Model 1 day care arrangements have an impact on the distribution of skills: this is the model studied in the previous sections. The set-up we label Model 2 differs from Model 1 only because day care arrangements do not have such an impact, that is $\pi_n^j(n_t^j) = 0 \ \forall j$. 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 find the parameters characterizing these functions using a calibration procedure. Once we have set these parameters, we compute in both models the optimal policies chosen by the government: the comparison of the optimal policies across the two models allow us to isolate the role played by the relationship between day care arrangements and the distribution of skills, which is the central point of our model. The existence of such a relationship is supported by the evidence mentioned in Section 1 and thus Model 1 is considered as the "true" model. Accordingly, the policies computed using Model 1 are denoted as "true optimal policies" and the policies computed using Model 2 are viewed as "mistaken optimal policies". Finally, we compute the welfare loss implied by the use in Model 1 of the mistaken optimal policies instead of the true optimal policies.

The numerical example is performed under the assumption of perfect correlation between market ability and ability to rear children. We focus our attention on the steady state and therefore we omit time subscripts. Regarding the fiscal instruments, we performed the numerical analysis both for the linear tax system described in Section 3 and for the mixed tax system of Section 4. However, in the presentation of the results, the focus is on the mixed tax system.[15] Moreover, we restrict our attention to the case where there is no laundering of the warm-glow component of individual preferences.

[15] The results for the linear tax system are available upon request.

6.1 Functional forms

We assume the following functional form for the utility function:

$$U^j = \alpha \frac{(c^j)^\xi}{\xi} + \beta^j \log z^j + \gamma^j \log n^j + \delta \left[\pi^j(n^j) \log H^2 + (1 - \pi^j(n^j)) \log H^1 \right], \tag{33}$$

with $\alpha + \beta^j + \gamma^j + \delta = 1$, $\alpha > 0$, $\beta^j > 0$, $\gamma^j > 0$, $\delta > 0$ and $\xi < 1$. Notice that, as we further explain in Section 6.2, the parameters for leisure time and for time with children, respectively β^j and γ^j, are allowed to be potentially type-specific; moreover we also allow for $\xi \neq 0$, that is the utility of consumption can be different from the logarithmic one.

The probability of becoming high skilled is:

$$\pi^j(n^j) = \iota^j + \sigma \frac{x^j}{1 + x^j}, \tag{34}$$

where x^j represents the quality of the early childhood environment of an agent of type j and it is produced combining parental time and day care services using the following constant elasticity of substitution (CES) function:

$$x^j = \left[\left(n^j H^j\right)^{\frac{1}{\nu}} + \left(d^j e\right)^{\frac{1}{\nu}} \right]^\nu, \tag{35}$$

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

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

$$p = \left(\frac{e}{\omega}\right)^{\frac{1}{\kappa}}, \tag{36}$$

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

6.2 Parameterization and Calibration

We interpret each period as having a length of 25 years. We set $\bar{a} = 0.2$, which means that over a period of 25 years, 5 years are spent receiving child care. In equation (4) we normalize A to 1.

Allocation of time

We classify as type 1-agents all the individuals with no college degree, while agents of type 2 are those with college education and above. We normalize to 1 the market productivity of agents of type 1: $H^1 = 1$. We set $H^2 = 2.3$ which is consistent with the data of Restuccia and Urrutia (2004).

We want to choose the parameters in the utility function $(\xi, \alpha, \beta^j, \gamma^j, \delta)$ in order to generate a realistic allocation of time between labor, parental time with the children and leisure for the two groups of agents. For this purpose we consider average data coming from the Harmonized

European Time Use Survey (HETUS).[16] Assuming, as it is usually done (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: $l^1 = 30\% < l^2 = 35\%$, $n^1 = 6\% < n^2 = 7\%$, $z^1 = 64\% > z^2 = 58\%$.[17] It should be stressed that high-skilled agents devote more time to their kids, though they also work more (see also Guryan et al. 2008). Notice that in Model 1 the value of δ affects the allocation of time since parental time with kids n has an indirect impact on the utility function through the probability of being a high skilled child π^j, which is type-specific. As a consequence we do not need to assume heterogeneous preferences to match the allocation of time of the two groups and thus we assume $\beta^j = \beta$ and $\gamma^j = \gamma$. Indeed we have four parameters ($\xi, \alpha, \beta, \delta$, with γ residually determined by the restriction $\alpha+\beta+\gamma+\delta = 1$) to match four target variable (n^1, n^2, l^1, l^2 with z^j residually determined by the time constraint (2)). In Model 2, the parameter δ does not affect the allocation of time since the probability of becoming high-skilled π^j does not depend on parental time n: for this reason we normalize δ to 0. Thus, to match the allocation of time of the two groups we need to allow for different β^j (and therefore different γ^j, with γ^j residually determined by the restrictions $\alpha + \beta^j + \gamma^j = 1$), in order to have once again four parameters ($\xi, \alpha, \beta^1, \beta^2$) to match the four target variables characterizing the allocation of time.

Child care and human capital

As to the parameter ν, which affects the elasticity of substitution between parental time and day care services in the production of the quality of the early childhood environment x, we are aware of the existing estimates referring to the elasticity of substitution between inputs in the production of the general category of home produced goods. For example, the estimates of Aguiar and Hurst (2007) would suggest a value of $\nu = 0.6$. However, these estimates refer to a large set of home produced goods rather than to the quality of the early childhood environment determined by child care. Direct estimates of the elasticity of substitution between non-parental time and parental time in the production of the quality of the early childhood environment are not available to the best of our knowledge. The degree of substitutability between inputs may differ when the focus is not on the general category "home production" but rather on the output of child care choices. Rogerson (2007), for instance, suggests that in the latter case the elasticity of substitution may be higher: he uses $\nu = 0.8$. We choose the same value.

As to the elasticity κ of the quality of day care services with respect to their cost, there is no direct evidence (to the best of our knowledge). In our benchmark simulation we choose $\kappa = 0.5$. In Appendix D we perform a sensitivity analysis on this parameter.

[16] 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.

[17] Two remarks are important in interpreting these data. First, parental time with children 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.

The parameter ω in equation (36) is chosen so that the ratio $p/(wH^1)$ between the cost of day care services and the wage of the unskilled agents is equal to 50%, which is consistent with the evidence in Blomquist et al. (2010).

The parameters of the probability function (34) ι^1 and ι^2 are chosen to match the probabilities of becoming a high-skilled agent: $\pi^1 = 0.23$, $\pi^2 = 0.65$ (see Caucutt et al. 2003), which imply $f^1 = 0.6$ and $f^2 = 0.4$.

In Model 1, the parameter σ of the probability function (34) and the quality of day care services e are chosen to match specific values of the elasticities ψ^j of the probabilities of becoming a high skilled with respect to parental time, with $\psi^j \equiv \frac{\pi_n^j(n^j)}{\pi^j(n^j)} n^j$. Our benchmark choices are $\psi^1 = -0.04$ and $\psi^2 = 0.04$, meaning that a 10% increase in the time low (high) skilled parents devote to their kids reduces (increases) their probability to become high skilled agents by 0.4% . In Appendix D we perform a sensitivity analysis and consider different values for the elasticities ψ^j.

To put these number on the elasticities ψ^j in perspective, we can refer to the empirical analysis of Bernal and Keane (2009). In their baseline specification they find that, on average, a year of day care translates into a 0.040 to 0.053 year reduction in completed schooling. As suggested by Bernal and Keane (2009) themselves, it is useful to look at the implication of these estimates for the rate of change in the proportion of people with a college degree. Assume, as in our model, that people are of two types, those who finish high school (12 years of education) and those who finish college (16 years of education), and that 40% of agents finish college. To increase average completed schooling by 0.040 years, the percentage finishing college must increase to 41%. Thus, following the empirical evidence of Bernal and Keane (2009), a one year reduction in the use of day care services implies on average a 2.5% increase in the proportion of people with a college degree. The values of ψ^1 and ψ^2 we use in our benchmark simulation, imply that in steady state a one year reduction in the use of day care by both high skilled and low skilled parents increases the proportion of people with a college degree by 1%.[18]. Therefore the strength of the relationship between child care and the distribution of skills implied by our choice of ψ^1 and ψ^2 does not seem to be too high compared to the available evidence. Nonetheless, we find in Section 6.3 that disregarding the effects of parental time on children's human capital in the design of the optimal policy entails a sizable welfare loss.

The quality of day care services e, determined according to the above procedure, will be called "calibrated quality". Notice that, when we simulate the optimal taxation problem in the next section, we allow the social planner to optimally choose a level of quality of day care services which differs from the calibrated one. Analogously, notice that the above specified

[18] In steady state, it is possible to show that the rate of change of the probability of becoming high skilled implied by a one year reduction in the use of day care by both high skilled and low skilled parents is equal to: $\frac{1-\pi^2(n^2)}{1-\pi^2(n^2)+\pi^1(n^1)} \frac{\psi^1}{n^1} \frac{1}{25} + \frac{\pi^2(n^2)}{1-\pi^2(n^2)+\pi^1(n^1)} \frac{\psi^2}{n^2} \frac{1}{25}$ (recall that we interpret each period in the model as having a length of 25 years). The chosen value of ψ^1 implies that a one year reduction in the use of day care by low skilled parents ($\frac{1-\pi^2(n^2)}{1-\pi^2(n^2)+\pi^1(n^1)} \frac{\psi^1}{n^1} \frac{1}{25}$) decreases the fraction of people with a college degree f^2 by 1.6%. The chosen value of ψ^2 implies that a one year reduction in the use of day care by high skilled parents ($\frac{\pi^2(n^2)}{1-\pi^2(n^2)+\pi^1(n^1)} \frac{\psi^2}{n^2} \frac{1}{25}$) increases f^2 by 2.6%.

values of ψ^j reflect assumptions on the current situation and they may have different values once the optimal policies are calculated and implemented.

In Model 2, parental time has no effect on the child's probability of becoming a high skilled agents, that is $\sigma = 0$. Moreover, once the cost of producing child care p has been set (see above), the quality of day care services e does not play any role.

Fiscal variables

As discussed above, the calibration procedure requires that some parameters are chosen in order to make some variables of the model (namely, those concerning the allocation of time) equal to their empirically observed counterparts. The individual choices we observe in the data obviously depend on the tax/transfer scheme which is actually in place. Therefore to properly perform the calibration of the model we need to make assumptions on the current fiscal system.

We approximate the current income tax schedule using the following simple parametric tax function (see Li and Sarte 2004):

$$T^j = \chi \cdot (Y^j)^{1+\phi}, \tag{37}$$

with $Y^j = wH^j l^j$.

We choose χ 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). We choose $\phi = 0.6$, as in Prescott (2004).

We also have an *ad valorem tax* on day care, whose rate is τ^d, and an *ad valorem* tax on consumption, whose rate is τ^c. We allow the government to also use a lump-sum transfer G. We set the tax rate on day care τ_d equal to -90% which is consistent with the information given for Sweden by Blomquist et al. (2010). Finally, we choose the tax rate on consumption τ_c equal to 25% (see McDaniel 2007 tax data series). The lump-sum transfer is residually determined in order to a have a balanced government budget.

Notice that the (parametric) fiscal system described above is just used for the purpose of calibrating the model: the social planner will choose a fully non linear income tax.

Table 1 summarizes the parameters' values used in the benchmark simulation.

6.3 Simulation: optimal policy and welfare evaluation

Once we have determined the parameters of Model 1 and Model 2 using the procedure described above, we compute for both models the optimal policies. We focus on the mixed tax system discussed in Section 4, where the government chooses a potentially nonlinear income tax schedule and a linear tax on day care. As in Section 4, when determining the optimal fiscal policy, the tax rate on consumption τ_c can be normalized to 0. We assume that the discount rate of the planner ρ is equal to 0.9.

Table 2 shows the optimal policies. The first column reports the results of the simulation performed in Model 1 (where the skills' transmission endogenously depends on the time parents devote to their kids): in this case the quality is optimally chosen by the government. The second column shows the results of the simulation performed in Model 2 (where the process of

Table 1: Parameterization and Calibration of Model 1 and Model 2: Benchmark case

Care time	$\bar{a}$	$= 0.2$ (meaning that 20% of the first period is spent receiving child care)
Total factor productivity	A	normalized to 1
Human capital levels	H^1	normalized to 1
	H^2	$= 2.3$
Parameters of the utility function	$\xi, \alpha, \beta^j, \gamma^j, \delta$	chosen to match the allocation of time during parenthood between
		labor ($l^1 = 30\%$, $l^2 = 35\%$), parental time devoted to children ($n^1 = 6\%$,$n^2 = 6\%$) and leisure ($z^1 = 62\%$, $z^2 = 62\%$).
		Notice that in Model 1 $\beta^j = \beta$ and $\gamma^j = \gamma$ and in Model 2 δ is normalized to 0.
Parameters of the probability functions	ι^1 and ι^2	chosen to match the following probabilities of becoming high skilled: $\pi^1 = 0.23$, $\pi^2 = 0.65$
Parameters of the probability functions and quality of day care services	σ and e	in Model 1 chosen to have $\psi^1 = -0.04$ and $\psi^2 = 0.04$
		where $\psi^j \equiv \frac{\pi^j_n(n^j)}{\pi^j(n^j)} n^j$ is the elasticity of the probability of becoming a high skilled with respect to parental time
		in Model 2 $\sigma = 0$ and e does not play any role (once p has been set -see below.)
Parameter affecting the elasticity of substitution between parental time and day care services	ν	0.8
Elasticity of the quality of day care services with respect to their cost	κ	0.5
Scale parameter affecting the unitary cost of producing day care	ω	chosen to have a ratio between the unitary cost of day care services and the wage rate of the unskilled equal to 50%
Parameter of the income tax function	χ	chosen to have an average tax burden equal to 35%
Parameter of the income tax function	ϕ	$= 0.6$
Ad valorem tax rate on day care services	τ_d	$= -90\%$
Ad valorem tax rate on consumption	τ_c	$= 25\%$

Table 2: Optimal policies: benchmark case.

	Model 1	Model 2
$\bar{T}^1$	-0.397	-0.467
$\bar{T}^2$	0.239	0.228
T'^1	0.027	0.064
T'^2	0.067	0.005
τ_d	-0.412	-0.157
$METR^1$	-0.029	0.045
$METR^2$	0.048	0.000
e	2.074	1.715
$\theta - 1$	0.019	
ψ^1	-0.056	0.000
ψ^2	0.020	0.000

Table 3: Values of the main variables at the optimum: benchmark case.

	Model 1	Model 2
$\pi^1(n)$	0.260	0.230
$\pi^2(n)$	0.659	0.650
f^1	0.568	0.603
f^2	0.432	0.397
n^1	0.057	0.065
n^2	0.026	0.027
d^1	0.183	0.175
d^2	0.214	0.213
l^1	0.557	0.519
l^2	0.777	0.778

skills' transmission is entirely exogenous). In this case the optimal problem of the government does not concern the choice of the quality of day care services, which is taken as given and fixed at the value found by the calibration procedure described above. The first eight rows of the Table report the optimal policy, that is: the average and marginal tax rates on income $\bar{T}^j$ and T'^j, the tax rate on day care services τ_d, the marginal effective tax rate $METR^j$ and the optimal value of the quality of day care services e. The existence of a relationship between the distribution of skills and child care arrangements is supported by the evidence mentioned in Section 1 and thus Model 1 is considered as the "true" model. Accordingly, the policies computed using Model 1 are denoted as "true optimal policies" and the policies computed using Model 2 are viewed as "mistaken optimal policies". The ninth row of Table 2 computes the welfare loss implied by the use in Model 1 of the mistaken optimal policies instead of the true optimal policies.[19] We measure this welfare loss as the percentage increase in consumption of both types of agents needed to leave social welfare unchanged. In other terms we compute the factor θ which, when multiplied by the consumption of all agents in the equilibrium allocation where the mistaken optimal policies are used, achieves the same level of social welfare as under the equilibrium allocation where the true optimal policies are used. The welfare loss measure is then obtained as $\theta - 1$. Finally, in the last two rows of the Table we show the values - computed at the optimum - of the elasticities ψ^j of the probabilities of becoming a high skilled with respect to parental time. Indeed, as explained in Section 4, the sign of the impact of parental time n on the child's probability π^j of becoming a high skilled type has an important effect on the design of the optimal policy as one can appreciate comparing across Model 1 and Model 2 the results reported in the rest of the Table. In Table 3 we also report for the sake of completeness the values at the optimum of the other main variables of the model.

To understand why the optimal policy is different across Model 1 and Model 2, it is useful to recall from Section 4 that a distinctive feature of Model 1 is the presence of a type-specific Pigouvian externality: due to the warm-glow altruism, parents do not fully take into account the impact of the chosen child care arrangement on the utility of their kids; moreover, such an impact may be different for the high skilled and for the low skilled. As we can see from Table 2, at the optimum $\psi^1 < 0$ and $\psi^2 > 0$, meaning that parental time has a negative (positive) effect on the probability of being high skilled for the kids of the low (high) skilled agents. Therefore the tax system should induce the low (high) skilled to devote less (more) time to their kids.

As a consequence the marginal tax rates on income for the low skilled is lower in Model 1 than in Model 2 (2.7% vs. 6.4%). Analogously, the marginal tax rate on income for the high skilled is higher in Model 1 than in Model 2 (6.7% vs. 0.5%). A similar remark applies to the marginal effective tax rates on income (see Section 4 for the definition of this measure).

[19] We check that the self-selection constraint is verified also when the mistaken policy are used in Model 1. Notice that when the mistaken policies are implemented in Model 1, the government budget constraint is no longer satisfied (the government runs a deficit). We allow the government to use a lump-sum tax in order to keep its budget balanced.

In particular, notice that $METR^2$ is equal to 0 in Model 2 (the standard no-distortion at the top result), while it is positive (and equal to 4.8%) in Model 1. $METR^1$ is positive in Model 2 (4.5%) and negative in Model 1 (−2.9%).

As to the linear tax rate on day care, it has to average the adjustments ideally required to correct the behavior of the two types of agents. As we can see from Table 3, the *ad valorem* tax rate on day care is more negative in Model 1 than in Model 2 (−41.2% vs. −15.7%), meaning that the subsidy paid by the government is higher when the link between child care and child development is considered: the need to encourage the low skilled to use day care services prevail on the need to increase parental time provided to kids by the high skilled.

Finally, the optimal quality of day care services turn out to be greater than the calibrated market quality (2.074 vs. 1.715). This implies that the values of ψ^1 and ψ^2 at the optimum (respectively −0.056 and 0.020) are lower than the values of ψ^1 and ψ^2 assumed in the calibration procedure (−0.04 and 0.04). Therefore at the optimum the use of day care services has a more positive (less negative) effect on children of low (high) skilled parents.

The higher quality of day care services, along with the higher value of the subsidy to day care, explains why both low skilled parents and high skilled parents use more day care services in Model 1 than in Model 2 (see Table 3).

Notice that the difference between the policies computed using Model 1 (the true optimal policies) and those computed under Model 2 (the mistaken optimal policies) is remarkable. The welfare loss $\theta - 1$, which gives information on the importance of using the true optimal policies instead of the mistaken optimal policies, is also sizable. The consumption of all agents need to be increased by 1.9% in the equilibrium allocation determined by the mistaken optimal policies in order to achieve the same level of social welfare as under the equilibrium allocation achieved under the true optimal policies. In Appendix D we perform a sensitivity analysis on some parameters of the model: we show that the welfare loss may range from 0.2% to 5.7%.

7 Summary and Discussion

This paper has characterized the optimal tax policy and the optimal level of quality of day care in a OLG model where parental choices over child care arrangements affect in a type-specific way the probability that a child becomes a high-market-ability adult. Two different scenarios have been considered: first, one where the government's tax instruments are restricted to be linear; second, a set-up where labor income is allowed to be taxed nonlinearly. In both cases we have discussed the rules dictating the optimal choice of day care quality enforced by the government.

With respect to previous contributions, optimal tax formulas incorporate type-specific Pigouvian terms which correct for the intergenerational externality in the human capital accumulation process. 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 expenditures that is affected for externality-correcting purposes; all tax formulas are affected. As far as the optimal choice of the quality of day care is concerned, this is determined by equating the total private marginal benefits of a quality increase to the marginal cost adjusted for the presence of three additional terms: the first one captures the impact on the government's budget constraint through the change in demand of consumption and day care services. The second one reflects the intergenerational externality in human capital accumulation. The third one, which is present only in the mixed tax system, comes from the self-selection constraint.

Finally, to get insights on the practical relevance of the mechanisms investigated in our theoretical model, we have supplemented our analysis with some numerical simulations. The results that we have obtained indicate that a public policy that disregards the effects of parental time on children's human capital entails a welfare loss that ranges from 0.2% to 5.7% of aggregate consumption.

An assumption that we have maintained throughout the paper is that the government's objective function is given by the discounted sum of agents' utilities; the only deviation from a purely utilitarian setting was the possibility to launder, fully or partially, the warm-glow component of an agent's utility in the social welfare function. This raises the question of how results would change under 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. To answer this question, suppose that the government assigns a welfare weight α^i to agents of type i. Then, the only effect on the tax- and quality-of-day-care-formulas would be 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 to choosing a relatively small (resp.: large) value of ε for that type of agent in the social welfare function.

Appendix

Appendix A

DERIVATION OF (15) AND (16)

The first order condition with respect to G_t is:

$$\begin{aligned}
&\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left[\varkappa_t^{jk} - (1-\varepsilon)\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial G_t}(\eta(H^2)-\eta(H^1))\right] + \\
&\mu_t\left[\tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial c_t^{jk}}{\partial G_t} + \tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial d_t^{jk}}{\partial G_t} - 1\right] + \\
&\upsilon_t\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial G_t} \\
= \;& 0. \qquad (38)
\end{aligned}$$

Using (14) it is immediate to see that (38) can be rewritten as (15).

The first order conditions with respect to τ_t^x $(x = c, d)$ is:

$$\begin{aligned}
&\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left[-\varkappa_t^{jk}x_t^{jk} - (1-\varepsilon)\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial \tau_t^x}(\eta(H^2)-\eta(H^1))\right] + \\
&\mu_t\left[\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}x_t^{jk} + \tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial c_t^{jk}}{\partial \tau_t^x} + \tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial d_t^{jk}}{\partial \tau_t^x}\right] + \\
&\upsilon_t\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial \tau_t^x} = 0. \qquad (39)
\end{aligned}$$

Using the Slutsky equation and rearranging terms we can rewrite (39) as (16).

DERIVATION OF (17)

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

$$\begin{aligned}
&\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left[\frac{\partial V_t^{jk}}{\partial e_t} - (1-\varepsilon)\left(\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial e_t} + \frac{\partial \pi^{jk}}{\partial e_t}\right)(\eta(H^2)-\eta(H^1))\right] + \\
&\mu_t\left(\tau_t^c\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial c_t^{jk}}{\partial e_t} + \tau_t^d\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial d_t^{jk}}{\partial e_t}\right) + \\
&\upsilon_t\sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left(\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial e_t} + \frac{\partial \pi^{jk}}{\partial e_t}\right) + p'(e_t)\,\Upsilon_t \\
= \;& 0, \qquad (40)
\end{aligned}$$

where Υ_t has been defined as:

$$\begin{aligned}\Upsilon_t \quad \equiv \quad & \sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\left[\frac{\partial V_t^{jk}}{\partial q_t} - (1-\varepsilon)\left(\eta(H^2)-\eta(H^1)\right)\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial q_t}\right] + \\ & \mu_t\left(\tau_t^c \sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial c_t^{jk}}{\partial q_t} + \tau_t^d \sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial d_t^{jk}}{\partial q_t}\right) + \\ & \upsilon_t \sum_{j=1}^{2}\sum_{k=1}^{2} f_t^{jk}\frac{\partial \pi^{jk}}{\partial d_t^{jk}}\frac{\partial d_t^{jk}}{\partial q_t},\end{aligned} \tag{41}$$

with $q_t \equiv p(e_t) + \tau_t^d$ indicating the consumer price for one hour of day-care. Using the first order condition with respect to τ_t^d (39), 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 services. We can therefore rewrite (40) as (17).

Appendix B

DERIVATION OF (22) AND (23)

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

$$\left(f_t^2+\lambda_t\right)\frac{\partial V_t^2}{\partial Y_t^2} \quad = \quad \left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial Y_t^2} - \mu_t\left(1+\tau_t\frac{\partial d_t^2}{\partial Y_t^2}\right)\right\} f_t^2; \tag{42}$$

$$\left(f_t^2+\lambda_t\right)\frac{\partial V_t^2}{\partial B_t^2} \quad = \quad \left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2} - \mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right\} f_t^2. \tag{43}$$

Dividing (42) by (43) and multiplying the result by the right hand side of (43), we get:

$$\begin{aligned} & \frac{\frac{\partial V_t^2}{\partial Y_t^2}}{\frac{\partial V_t^2}{\partial B_t^2}}\left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2} - \mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right\} \\ = \quad & \left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial Y_t^2} - \mu_t\left(1+\tau_t\frac{\partial d_t^2}{\partial Y_t^2}\right). \end{aligned} \tag{44}$$

Using (20) to collect terms in (44) gives:

$$T'\left(Y_t^2\right) = -\left(\frac{dd_t^2}{dY_t^2}\right)_{dV_t^2=0}\tau_t + \frac{1}{\mu_t}\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^2}{\partial n_t^2}\left(\frac{dn_t^2}{dY_t^2}\right)_{dV_t^2=0}. \tag{45}$$

We can then use (45) to substitute terms in (21) and in this way obtain (22).

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

$$f_t^1\frac{\partial V_t^1}{\partial Y_t^1} = \lambda_t\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}+\left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial Y_t^1}-\mu_t\left(1+\tau_t\frac{\partial d_t^1}{\partial Y_t^1}\right)\right\}f_t^1; \tag{46}$$

$$f_t^1\frac{\partial V_t^1}{\partial B_t^1} = \lambda_t\frac{\partial \widehat{V}_t^2}{\partial B_t^1}+\left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1}-\mu_t\left(-1+\tau_t\frac{\partial d_t^1}{\partial B_t^1}\right)\right\}f_t^1. \tag{47}$$

Dividing (46) by (47) and multiplying the result by the right hand side of (47), we get:

$$\begin{aligned}&\frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\left\{\lambda_t\frac{\partial \widehat{V}_t^2}{\partial B_t^1}+\left[\left((1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right)\frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1}-\mu_t\left(-1+\tau_t\frac{\partial d_t^1}{\partial B_t^1}\right)\right]f_t^1\right\}\\ =\;&\lambda_t\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}+\left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial Y_t^1}-\mu_t\left(1+\tau_t\frac{\partial d_t^1}{\partial Y_t^1}\right)\right\}f_t^1.\end{aligned} \tag{48}$$

Using (20) to collect terms in terms in (48) gives:

$$\begin{aligned}T'\left(Y_t^1\right) &= -\left(\frac{dd_t^1}{dY_t^1}\right)_{dV_t^1=0}\tau_t+\frac{\lambda_t}{\mu_t f_t^1}\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}}-\frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\right)+\\ &\quad\frac{1}{\mu_t}\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^1}{\partial n_t^1}\left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0}.\end{aligned} \tag{49}$$

Using (49) to substitute terms in (21) gives (23).

DERIVATION OF (24)

The first order condition for τ_t is given by:

$$\begin{aligned}&\sum_{j=1}^{2}\left[\frac{\partial V_t^j}{\partial \tau_t}-(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial \tau_t}\right]f_t^j+\\ &\mu_t\sum_{j=1}^{2}\left(d_t^j+\tau_t\frac{\partial d_t^j}{\partial \tau_t}\right)f_t^j+\lambda_t\left(\frac{\partial V_t^2}{\partial \tau_t}-\frac{\partial \widehat{V}_t^2}{\partial \tau_t}\right)\\ =\;& -\upsilon_t\sum_{j=1}^{2}f_t^j\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial \tau_t}.\end{aligned} \tag{50}$$

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:

$$\begin{aligned}&\sum_{j=1}^{2}\left[-d_t^j\frac{\partial V_t^j}{\partial B_t^j}-(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial \tau_t}\right]f_t^j+\\ &\mu_t\sum_{j=1}^{2}\left(d_t^j+\tau_t\frac{\partial d_t^j}{\partial \tau_t}\right)f_t^j+\lambda_t\left(-d_t^2\frac{\partial V_t^2}{\partial B_t^2}+\widehat{d}_t^2\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\right)\\ =\;& -\upsilon_t\sum_{j=1}^{2}f_t^j\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial \tau_t}.\end{aligned} \tag{51}$$

Multiplying (43) and (47) 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 $-d_t^1 \frac{\partial V_t^1}{\partial B_t^1}$:

$$-\left(f_t^2+\lambda_t\right)\frac{\partial V_t^2}{\partial B_t^2} d_t^2 = -\left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right\} f_t^2 d_t^2; \tag{52}$$

$$-f_t^1 d_t^1\frac{\partial V_t^1}{\partial B_t^1} = -\lambda_t d_t^1\frac{\partial \widehat{V}_t^2}{\partial B_t^1}-\left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1}-\mu_t\left(-1+\tau_t\frac{\partial d_t^1}{\partial B_t^1}\right)\right\} f_t^1 d_t^1. \tag{53}$$

Substituting (52) and (53) into (51) 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:

$$\begin{aligned}
&-\left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right\} f_t^2 d_t^2 + \\
&-\lambda_t d_t^1\frac{\partial \widehat{V}_t^2}{\partial B_t^1}-\left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1}-\mu_t\left(-1+\tau_t\frac{\partial d_t^1}{\partial B_t^1}\right)\right\} f_t^1 d_t^1 + \\
&\mu_t\sum_{j=1}^{2}\left[d_t^j+\tau_t\left(\frac{\partial \widetilde{d}_t^j}{\partial \tau_t}-d_t^j\frac{\partial d_t^j}{\partial B_t^j}\right)\right] f_t^j-\sum_{j=1}^{2}(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial \tau_t} f_t^j+\lambda_t \widehat{d}_t^2\frac{\partial \widehat{V}_t^2}{\partial B_t^1} \\
= \; &-\upsilon_t\sum_{j=1}^{2} f_t^j\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial \tau_t},
\end{aligned} \tag{54}$$

which can be simplified to obtain:

$$\begin{aligned}
\sum_{j=1}^{2} f_t^j\tau_t\frac{\partial \widetilde{d}_t^j}{\partial \tau_t} &= -\frac{\upsilon_t}{\mu_t}\left[\sum_{j=1}^{2} f_t^j\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial \tau_t}+\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2} f_t^2 d_t^2+\frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1} f_t^1 d_t^1\right] + \\
&\frac{(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)}{\mu_t}\sum_{j=1}^{2} f_t^j\frac{\partial \pi^j}{\partial n_t^j}\left(\frac{\partial n_t^j}{\partial \tau_t}+\frac{\partial n_t^j}{\partial B_t^j} d_t^j\right)+\frac{\lambda_t}{\mu_t}\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(d_t^1-\widehat{d}_t^2\right).
\end{aligned} \tag{55}$$

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 (55) in a more compact form as:

$$\sum_{j=1}^{2} f_t^j\tau_t\frac{\partial \widetilde{d}_t^j}{\partial \tau_t} = \frac{(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)}{\mu_t}\sum_{j=1}^{2} f_t^j\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial \widetilde{n}_t^j}{\partial \tau_t}+\frac{\lambda_t\frac{\partial \widehat{V}_t^2}{\partial B_t^1}}{\mu_t}\left(d_t^1-\widehat{d}_t^2\right)-\frac{\upsilon_t}{\mu_t}\sum_{j=1}^{2} f_t^j\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial \widetilde{n}_t^j}{\partial \tau_t}.$$

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 obtain (24).

DERIVATION OF (27)

The first order condition with respect to e_t is given by:

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

where, using the notation $q_t \equiv p\left(e_t\right)+\tau_t$, Υ_t has been defined as:

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

The term $p'\left(e_t\right)\Upsilon_t$ captures the effects of the increase in the unitary price of day care services (due to the higher quality level) on the Lagrangian of the government's problem. 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 first order condition with respect to e_t as:

$$\begin{aligned} & \left(f_t^1\frac{\partial V_t^1}{\partial B_t^1}-\lambda_t\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\right)\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}}+\left(f_t^2+\lambda_t\right)\frac{\partial V_t^2}{\partial B_t^2}\frac{\frac{\partial V_t^2}{\partial e_t}}{\frac{\partial V_t^2}{\partial B_t^j}}+\lambda_t\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}}-\frac{\frac{\partial \widehat{V}_t^2}{\partial e_t}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}}\right)+ \\ & \sum_{j=1}^{2}\left[-(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)\left(\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial e_t}+\frac{\partial \pi^j}{\partial e_t}\right)\right]f_t^j+\mu_t\sum_{j=1}^{2}\tau_t\frac{\partial d_t^j}{\partial e_t}f_t^j \\ = & -\upsilon_t\sum_{j=1}^{2}\left(\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial e_t}+\frac{\partial \pi^j}{\partial e_t}\right)f_t^j-p'\left(e_t\right)\Upsilon_t. \end{aligned} \tag{56}$$

Now use (43) and (47) 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 (56). This gives:

$$
\begin{aligned}
&\left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial\pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1}-\mu_t\left(-1+\tau_t\frac{\partial d_t^1}{\partial B_t^1}\right)\right\}f_t^1\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}}+\\
&\left\{\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\frac{\partial\pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right\}f_t^2\frac{\frac{\partial V_t^2}{\partial e_t}}{\frac{\partial V_t^2}{\partial B_t^j}}+\mu_t\sum_{j=1}^{2}\tau_t\frac{\partial d_t^j}{\partial e_t}f_t^j+\\
&\lambda_t\frac{\partial\widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}}-\frac{\frac{\partial\widehat{V}_t^2}{\partial e_t}}{\frac{\partial\widehat{V}_t^2}{\partial B_t^1}}\right)-(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)\sum_{j=1}^{2}\left(\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial e_t}+\frac{\partial\pi^j}{\partial e_t}\right)f_t^j\\
=\;&-\upsilon_t\sum_{j=1}^{2}f_t^j\left(\frac{\partial\pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial e_t}+\frac{\partial\pi^j}{\partial e_t}\right)-p'\left(e_t\right)\Upsilon_t.
\end{aligned}
\tag{57}
$$

From (50) we can see that at an optimum it is $\Upsilon_t=-\mu_t\sum_{j=1}^{2}d_t^jf_t^j$. Therefore, dividing by μ_t all terms in the previous equation and rearranging terms, we get:

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

Using (28) and (29) we can express the condition implicitly defining the optimal level of day care quality as:

$$
\begin{aligned}
\sum_{j=1}^{2}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}f_t^j=&\frac{1}{\mu_t}\sum_{j=1}^{2}\left[(1-\varepsilon)\left(\eta\left(H^2\right)-\eta\left(H^1\right)\right)-\upsilon_t\right]\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0}f_t^j+\\
&\frac{\lambda_t}{\mu_t}\frac{\partial\widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial\widehat{V}_t^2}{\partial e_t}}{\frac{\partial\widehat{V}_t^2}{\partial B_t^1}}-\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}}\right)-\tau_t\sum_{j=1}^{2}\left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0}f_t^j+p'\left(e_t\right)\sum_{j=1}^{2}d_t^jf_t^j.
\end{aligned}
\tag{58}
$$

Finally, using (26), we can rewrite (58) as (27).

Appendix C

AGENT-MONOTONICITY ASSUMPTION

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 j.[20] 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 $u\left(c^j, z^j, n^j\right) + \pi^j(n^j)\eta(H^2) + (1-\pi^j(n^j))\eta(H^1)$, the conditional indirect utility for an agent of type j, $V^j\left(Y, B\right)$, is obtained maximizing $u\left(c^j, z^j, n^j\right) + \pi^j(n^j)\eta(H^2) + (1-\pi^j(n^j))\eta(H^1)$ subject to the budget constraint $c^j = B - \left(p\left(e\right)+\tau\right)d^j = B - \left(p\left(e\right)+\tau\right)\left(l^j + z^j\right) = B - \left(p\left(e\right)+\tau\right)\left[\left(Y/wH^j\right) + z^j\right]$ 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 $\partial V^j/\partial Y = -\left(wH^j\right)^{-1}\left[\frac{\partial u\left(\cdot; wH^j\right)}{\partial c^j}\left(p\left(e\right)+\tau\right) + \frac{\partial u\left(\cdot; wH^j\right)}{\partial n^j} + \left(\eta(H^2) - \eta\left(H^1\right)\right)\frac{\partial \pi^j}{\partial n^j}\right]$. Thus, we have:

$$
\begin{aligned}
-\frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}} &= \left(wH^1\frac{\partial u\left(\cdot; wH^1\right)}{\partial c_t^1}\right)^{-1}\left[\frac{\partial u\left(\cdot; wH^1\right)}{\partial c_t^1}\left(p\left(e\right)+\tau_t\right) + \frac{\partial u\left(\cdot; wH^1\right)}{\partial n_t^1} + \left(\eta(H^2) - \eta\left(H^1\right)\right)\frac{\partial \pi^1}{\partial n_t^1}\right] \\
&= \frac{p\left(e\right)+\tau_t}{wH^1} + \frac{\frac{\partial u\left(\cdot; wH^1\right)}{\partial n_t^1} + \left(\eta(H^2) - \eta\left(H^1\right)\right)\frac{\partial \pi^1}{\partial n_t^1}}{wH^1\frac{\partial u\left(\cdot; wH^1\right)}{\partial c_t^1}}
\end{aligned}
\tag{59}
$$

$$
\begin{aligned}
-\frac{\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} &= \left(wH^2\frac{\partial \widehat{u}\left(\cdot; wH^2\right)}{\partial \widehat{c}_t^2}\right)^{-1}\left[\frac{\partial \widehat{u}\left(\cdot; wH^2\right)}{\partial \widehat{c}_t^2}\left(p\left(e\right)+\tau_t\right) + \frac{\partial \widehat{u}\left(\cdot; wH^2\right)}{\partial \widehat{n}_t^2} + \left(\eta(H^2) - \eta\left(H^1\right)\right)\frac{\partial \widehat{\pi}^2}{\partial \widehat{n}_t^2}\right] \\
&= \frac{p\left(e\right)+\tau_t}{wH^2} + \frac{\frac{\partial \widehat{u}\left(\cdot; wH^2\right)}{\partial \widehat{n}_t^2} + \left(\eta(H^2) - \eta\left(H^1\right)\right)\frac{\partial \widehat{\pi}^2}{\partial \widehat{n}_t^2}}{wH^2\frac{\partial \widehat{u}\left(\cdot; wH^2\right)}{\partial \widehat{c}_t^2}}.
\end{aligned}
\tag{60}
$$

Comparing (59) with (60), 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,[21] 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 (60)) 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 (59)). Moreover, as our assumptions imply that $\left(\eta(H^2) - \eta\left(H^1\right)\right)\partial\widehat{\pi}^2/\partial\widehat{n}_t^2 > 0$ whereas $\left(\eta(H^2) - \eta\left(H^1\right)\right)\partial\pi^1/\partial n_t^1 < 0$, this also

[20] We suppress for simplicity the time subscripts.

[21] 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).

contributes to raise the marginal rate of substitution of the mimicker relatively to that of a true low-skilled. On the other hand, 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 $\partial u\left(\cdot; wH^1\right)/\partial n_t^1$.

Appendix D

In this Appendix we perform a sensitivity analysis on some parameters of the model, namely the elasticity κ of the quality of day care services with respect to the cost (see equation (36)), the parameter σ (see equation 34) and the calibrated quality of day care services e (see equation (35)). These parameters are expected to play a role for the results presented in Section 6 but we do not have compelling evidence on their values.

Table 4 shows the sensitivity analysis on κ. The higher κ is, the higher is the optimal level of quality of day care chosen in Model 1 and therefore the welfare loss due to the use of the mistaken optimal policies. We only consider values of κ such that the optimal level of quality chosen by the planner is between the calibrated one (1.715) and the human capital of high skilled agents H^2 (2.3), which we can reasonably view as an upper bound for the quality of day care services. In particular, when $\kappa = 0.38$ the optimal level of quality almost coincides with the calibrated quality: in this case the welfare loss (0.3%) can be still considered remarkable but it is much smaller than the one reported in Table 2. When $\kappa = 0.55$, the optimal level of quality is almost equal to the human capital of the high skilled agents: in this case the welfare loss is very high (3%).

In Table 4, the parameter σ and the calibrated value of the quality of day care services e are still chosen as in the main text, that is they are set in such a way that $\psi^1 = -0.04$ and $\psi^2 = 0.04$, where $\psi^j \equiv \frac{\pi_n^j(n^j)}{\pi^j(n^j)}n^j$ is the elasticity of the probability of becoming a high skilled with respect to parental time. Tables 5-7 present the results of a sensitivity analysis in which we choose σ and e in order to have different values of the elasticities of the probabilities of becoming high skilled with respect to parental time, while still considering as in Table 4 an interval of values of κ such that the optimal quality of day care services is between the calibrated one and the human capital of the high skilled agents.[22]

In Tables 5 and 6 we still assume that the effect of parental time on the distribution of children's skills is of opposite sign for the unskilled and for the skilled. Namely, in the calibration procedure, we assume $\psi^1 = -0.01$ and $\psi^2 = 0.01$ in Table 5[23] and $\psi^1 = -0.07$ and $\psi^2 = 0.07$ in Table 6.[24]

[22] Notice that the values of κ which imply an optimal quality of day care services within this interval differ according to the values of σ and e. Moreover notice that, while the value of the human capital of the high skilled agents is always assumed to be equal to 2.3, the calibrated value of e may change depending on the assumptions on ψ^j.

[23] These values of ψ^j imply that in steady state a one year reduction in the use of day care by both low (high) skilled parents changes the proportion of people with a college degree by -0.40% (0.64%). The overall effect of a one year reduction by both types of parents is 0.24%.

[24] These values of ψ^j imply that in steady state a one year reduction in the use of day care by both low (high) skilled parents changes the proportion of people with a college degree by -2.8% (4.5%). The overall effect of a one year reduction by both types of parents is 1.7%.

Table 7 shows the results of the simulation for $\psi^1 = 0$ and $\psi^2 = 0.04$,[25] that is the time given by the low skilled has no impact on the distribution of children's skills, while the time given by the high skilled has still a positive impact.[26]

All in all, in Tables 5-7, the welfare loss ranges from 0.2% to 5.7% of aggregate consumption.

[25] These values of ψ^j imply that in steady state a one year reduction in the use of day care by both low (high) skilled parents changes the proportion of people with a college degree by 0 (2.6%). The overall effect of a one year reduction by both types of parents is 2.6%.

[26] We don't report the case where $\psi^1 = -0.05$ and $\psi^2 = 0$. Indeed it turns out that to have $\psi^2 = 0$ we need to assume a calibrated quality above the level of the human capital of the high skilled H^2, which we view as a reasonable upper bound for the quality of day care services.

Table 4: Optimal policies: different values of κ; σ and e set in the calibration procedure in order to have $\psi^1 = -0.04$ and $\psi^2 = 0.04$.

	$\kappa = 0.38$	$\kappa = 0.55$
$\bar{T}^1$	-0.398	-0.399
$\bar{T}^2$	0.239	0.239
T'^1	0.024	0.028
T'^2	0.070	0.065
τ_d	-0.273	-0.456
$METR^1$	-0.001	-0.043
$METR^2$	0.061	0.041
e	1.736	2.266
$\theta - 1$	0.003	0.030
ψ^1	-0.034	-0.066
ψ^2	0.027	0.017

Table 5: Optimal policies: different values of κ; σ and e set in the calibration procedure in order to have $\psi^1 = -0.01$ and $\psi^2 = 0.01$

	$\kappa = 0.8$	$\kappa = 0.95$
$\bar{T}^1$	-0.431	-0.423
$\bar{T}^2$	0.231	0.234
T'^1	0.040	0.044
T'^2	0.029	0.027
τ_d	-0.203	-0.289
$METR^1$	0.021	0.004
$METR^2$	0.022	0.015
e	1.726	2.266
$\theta - 1$	0.002	0.009
ψ^1	-0.010	-0.021
ψ^2	0.007	0.004

Table 6: Optimal policies: different values of κ; σ and e set in the calibration procedure in order to have $\psi^1 = -0.07$ and $\psi^2 = 0.07$

	$\kappa = 0.25$	$\kappa = 0.43$
$\bar{T}^1$	-0.375	-0.391
$\bar{T}^2$	0.243	0.236
T'^1	0.006	0.004
T'^2	0.108	0.099
τ_d	-0.284	-0.523
$METR^1$	-0.018	-0.093
$METR^2$	0.098	0.065
e	1.725	2.289
$\theta - 1$	0.005	0.057
ψ^1	-0.053	-0.101
ψ^2	0.051	0.029

Table 7: Optimal policies: different values of κ; σ and e set in the calibration procedure in order to have $\psi^1 = 0$ and $\psi^2 = 0.04$

	$\kappa = 0.57$	$\kappa = 0.84$
$\bar{T}^1$	-0.446	-0.440
$\bar{T}^2$	0.222	0.226
T'^1	0.030	0.039
T'^2	0.064	0.056
τ_d	-0.033	-0.367
$METR^1$	0.026	-0.059
$METR^2$	0.063	0.032
e	1.338	2.278
$\theta - 1$	0.002	0.037
ψ^1	-0.006	-0.063
ψ^2	0.025	0.011

References

Aguiar, M., Hurst, E., 2007. Life cycle prices and production. American Economic Review 97(5), 1533–1559.

Andreoni, J., 1989. Giving with impure altruism: applications to charity and ricardian equivalence. Journal of Political Economy 97 (6), 1447–58.

Andreoni, J., 2006. Philanthropy. In: Kolm, S., Ythier, J. M. (Eds.), Handbook on the Economics of Giving, Reciprocity and Altruism. Vol. 1. Elsevier, Ch. 18, pp. 1201–1269.

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

Bernal, R., Keane, M., 2009. Child care choices and children's cognitive achievement: the case of single mothers, mimeo.

Bernal, R., Keane, M., 2010. Quasi-structural estimation of a model of child-care choices and child cognitive ability production. Journal of Econometrics 156(1), 164–189.

Blomquist, S., Christiansen, V., Micheletto, L., 2010. Public provision of private goods and nondistortionary marginal tax rates. American Economic Journal: Economic Policy 2(2), 1–27.

Blomquist, S., Micheletto, L., 2009. Nonlinear income taxation and matching grants in a federation with decentralized in-kind transfers. International Economic Review 50 (2), 543–575.

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

Carneiro, P., Heckman, J., 2003. Human capital policy. In: Heckman, J., Krueger, A. (Eds.), Inequality in America: What Role for Human Capital Policies? MIT Press.

Carneiro, P., Meghir, C., Parey, M., 2007. Maternal education, home environments and the development of children and adolescents. Discussion Paper 3072, Institute for the Study of Labor (IZA).

Casarico, A., Sommacal, A., 2011. Labor income taxation, human capital and growth: the role of child care. The Scandinavian Journal of Economics (forthcoming).

Caucutt, E. M., Imrohoroglu, S., Kumar, K. B., 2003. Growth and welfare analysis of tax progressivity in a heterogeneous-agent model. Review of Economic Dynamics 6 (3), 546–577.

Cremer, H., Gahvari, F., 1997. In-kind transfers, self-selection and optimal tax policy. European Economic Review 41(1), 97–114.

Cremer, H., Pestieau, P., 2006. Intergenerational transfer of human capital and optimal education policy. Journal of Public Economic Theory 8, 529–545.

Cunha, F., Heckman, J., 2007. The technology of skill formation. American Economic Review 97, 31–47.

Cunha, F., Heckman, J., Lochner, L., Masterov, D., 2005. Interpreting the evidence on life cycle skill formation. In: Hanushek, E., Welch, F. (Eds.), Handbook of the Economics of Education. North Holland.

Diamond, P., 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.

Dixit, A., 1985. Tax policy in open economies. In: Auerbach, A. J., Feldstein, M. (Eds.), Handbook of Public Economics. Vol. 1. Elsevier, Ch. 6, pp. 313–374.

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

Galor, O., Zeira, J., 1993. Income distribution and macroeconomics. Review of Economics Studies 60, 35–52.

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

Glomm, G., Kaganovich, M., 2008. Social security, public education and the growth inequality relationship. European Economic Review 52, 1009–1034.

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

Guryan, J., Hurst, E., Kearney, M., 2008. Parental education and parental time with children. Journal of Economic Perspectives 22 (3), 23–46.

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

Havnes, T., Mogstad, M., 2010. Is universal child care leveling the playing field? evidence from non-linear difference-in-differences. Discussion Paper 4978, Institute for the Study of Labor (IZA).

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

Juster, F., 1985. A note on recent changes in time use. In: Juster, F., Stafford, F. (Eds.), Time, Goods, and Well-Being. Institute for Social Research, University of Michigan.

Kaplow, L., 1995. A note on subsidizing gifts. Journal of Public Economics 58 (3), 469–477.

Kaplow, L., 1998. Tax policy and gifts. American Economic Review 88 (2), 283–88.

Kimmel, J., Connelly, R., 2007. Mothers' time choices: Caregiving, leisure, home production, and paid work. Journal of Human Resources 42(3), 643–681.

Li, W., Sarte, P.-D., 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, mimeo.

Prescott, E., 2004. Why do americans work so much more than europeans? Federal Reserve Bank of Minneapolis Quarterly Review 28(1), 2–13.

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

Restuccia, D., Urrutia, C., 2004. Intergenerational persistence of earnings: the role of early and college education. American Economic Review 94, 1354–1378.

Rogerson, R., 2007. Taxation and market work: is scandinavia an outlier? Economic Theory 32 (1), 59–85.

Ruhm, C., 2004. Parental employment and child cognitive development. Journal of Human Resources 39, 155–192.

Schokkaert, E., 2006. The empirical analysis of transfer motives. In: Kolm, S.-C., Ythier, J. M. (Eds.), Handbook of the economics of giving, altruism and reciprocity. North-Holland, pp. 128–181.

Stiglitz, J. E., 1982. Self-selection and pareto efficient taxation. Journal of Public Economics 17 (2), 213–240.