Intergenerational transmission of skills during childhood and optimal public policy*

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

http://dx.doi.org/10.1007/s00148-014-0505-0

Abstract

We characterize the optimal tax policy and quality of day care services in an OLG model in which child care arrangements chosen by parents of different skill type affect the probability that children become high-skilled adults in a type-specific way. With respect to previous contributions, optimal tax formulas incorporate type-specific Pigouvian terms which correct for the intergenerational externality in human capital accumulation. The optimal quality of day care services is determined by equating the total private marginal benefits of a quality increase to its marginal cost, adjusted for the presence of three additional terms capturing respectively the budgetary impact of a change in demand for day care services, the intergenerational externality in human capital accumulation, and the self-selection constraint.

*We thank for insightful comments two anonymous referees as well as seminar participants at the SIEP conference in Pavia, the University of Louvain-la-Neuve, the University of St.Gallen, the University of Wien, the CESifo Area Conference on Employment and Social Protection and the IIPF Conference in Uppsala. All the remaining errors are ours.

[†]Bocconi University, Milan, Italy; CESifo, Munich, Germany. e-mail address: alessandra.casarico@unibocconi.it

[‡]Department of Law, University of Milan, Italy; Econpubblica, Bocconi University, Milan, Italy; Uppsala Center for Fiscal Studies, Sweden; CESifo, Munich, Germany. e-mail address: luca.micheletto@unimi.it

[§]Department of Economics, University of Verona, Italy; Econpubblica, Bocconi University, Milan, Italy. e-mail address: alessandro.sommacal@univr.it

KEYWORDS: optimal taxation; day care quality; intergenerational transmission of skills; early childhood environment; warm glow

JEL Classification: H21; H23; J13; J22; J24

1 Introduction

The paper aims at characterizing the optimal tax policy and quality of day care services in an OLG model in which child care arrangements chosen by parents of different skill type affect the probability that children become high-skilled adults in a type-specific way. By child care arrangements we mean parental time and time spent in day care centers.

The role of child care for 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 as a dynamic process, characterized by strong complementarities between early and late investments in human capital (Cunha et al. 2005; Cunha and Heckman 2007). The second one looks at the importance of parental time, and especially maternal time, vs. other types of child care in producing children abilities. Whereas the earlier contributions - as surveyed for instance by Ruhm (2004) - reached mixed conclusions, more recent research (see for instance Bernal 2008 and Bernal and Keane 2011, 2010) highlights that, on average, the substitution of maternal time with other child care sources produces negative and rather sizable effects on children's skills. However, this literature also shows that the sign of the impact depends on the type of child care that substitutes maternal time and on the level 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 are highly effective in promoting achievement.

Though the existing empirical literature supports the appropriateness of explicitly including child care in the skill formation process and it also indicates that the impact of parental time and day care services on the accumulation of human capital may be type-specific, the theoretical literature focusing on the design of public policies generally ignores these features. For instance, previous contributions in the optimal tax literature (see e.g. Balestrino 2000 and

[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.

Blomquist et al. 2010) have typically treated day care services as merely one prominent example of goods/services that are complements to labor and that, as such, should be subsidized in order to either mitigate the distortion on 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.[2]

The inclusion of child care in the skill formation process and the type-specific impact of child care arrangements on human capital accumulation will be the key aspects of our set-up. In our model parents derive utility from their own consumption, from leisure time with and without children, and from the children's human capital. The direct dependency of the parents' utility on the children's human capital represents a warm-glow component. We assume that agents are heterogeneous in human capital, whose distribution across individuals is endogenous, that is, it depends on parental choices over child care and on the parents' skill type.[3] The government is empowered with a nonlinear tax on labor income and a linear tax/subsidy on day care expenditures; in addition, it can enforce a level of day care quality which is optimally selected. We theoretically characterize the social welfare maximizing policy and also perform a numerical analysis to shed light on the quantitative relevance of the inclusion of child care in the human capital production function, both in terms of the optimal values of the policy variables and in terms of the welfare losses caused by setting policy neglecting its effect on child care arrangements and on the skill formation process.

The way parental altruism is specified is a crucial modelling issue in our model. There are several possible motivations behind child care decisions, including pure altruism (parents care about the utility of their children) and impure (warm-glow) altruism (parents derive joy from the level of human capital that child care arrangements deliver to their children). To the best of our knowledge, no direct empirical test exists on the type of altruism involved in the parental decision to devote time to children. However, the evidence on intrafamily income

[2] A model in which child care enters the human capital production function is analyzed by Casarico and Sommacal (2012) who study the impact on growth of changes in labor income taxation. However, their analysis is not concerned with the design of optimal public policies, which is instead the focus of the current paper.

[3] Heterogeneity in the ability to raise children is also considered in Balestrino et al. (2002, 2003). In these papers, however, the authors only allow for two possible uses of time, market labor supply and parental care, do not explicitly focus on external day care, and work within a static framework. On the other hand, they consider the possibility of endogenous fertility, which we neglect in our paper, and use weaker assumptions regarding the correlation between market ability and ability to raise children.

transfers, though not conclusive, rejects the predictions of pure altruism but tends to be consistent with warm-glow altruism (see Schokkaert (2006) for an exhaustive survey). Accordingly, we rely on the warm-glow assumption, which is shared by many papers on intergenerational transmission of human capital and wealth (e.g. Glomm and Ravikumar (1992); Glomm and Kaganovich (2008)) and it can also be found in the optimal taxation literature (e.g. Cremer and Pestieau (2006), Kopczuk (2013)).

Another important assumption in our model is that the quality of day care services is uniform and can be set optimally by the goverment. Of course this assumption better captures a situation in which day care is mainly publicly provided. This is not to disregard the role of private provision of day care services and the possibility for parents to select the provider that better matches their preferences. It is rather a first attempt to formally analyze the influence of day care quality on human capital accumulation and it indirectly captures the setting of (minimum) quality standards by the government, which is pervasive even in countries in which private provision is dominant.

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 and the evolution over time of the skill distribution in the population. Section 3 characterizes the optimal public policy. Section 4 provides a numerical analysis. Section 5 offers concluding remarks.

2 The model

2.1 The consumers

We consider a two-period OLG model with intragenerational heterogeneity: agents differ in their skill level H^j, which can only take two values, i.e. $j = 1, 2$, with $H^2 > H^1$. In the first period, agents (children) do not take any active choice; depending on child care arrangements and on the skill type of their parents, they have a certain probability to become high or low skilled. In the second period, each agent learns his own skill type, has a child, and decides how to allocate time between labor, leisure time with children (i.e. parental time devoted to children) and leisure time without children. Denoting time indices by a subscript, a parent of skill type j (hereafter labelled simply as parent of type j) maximizes the following expected utility function:

$$E(U_t^j) = \pi^j(n_t^j)\left[u\left(c_t^j, z_t^j, n_t^j\right) + \eta(H^2)\right] + (1-\pi^j(n_t^j))\left[u\left(c_t^j, z_t^j, n_t^j\right) + \eta(H^1)\right]$$
$$= u\left(c_t^j, z_t^j, n_t^j\right) + \pi^j(n_t^j)\eta(H^2) + (1-\pi^j(n_t^j))\eta(H^1), \qquad (1)$$

with $u''(\cdot) < 0 < u'(\cdot)$. We denote by c_t^j, z_t^j and n_t^j respectively consumption, leisure time without children and leisure time with children by a j-type agent. The underlying assumption is that leisure time with and without children are imperfect substitutes. The term $\eta(H^j)$ reflects the warm-glow altruism of parents towards the level of human capital of their children H^j, $j = 1, 2$, and $\eta'(\cdot) > 0$. The term $\pi^j(n_t^j)$ is the probability of having a high-skilled child and it stands for $\pi(n_t^j, H^j, e_t)$: the probability of being a high human capital agent is a function of the time n_t^j parents dedicate to children, the parents' type j, 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. High (resp.: low)-skilled parents are assumed to have, other things being equal, a higher (resp.: lower) probability to raise a child who will become a high-skilled adult. The implicit underlying assumption is that there is perfect correlation between human capital and ability to raise children (where ability to raise children is meant to capture ability to turn them into high-skilled adults).[4]

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

$$1 = l_t^j + n_t^j + z_t^j, \qquad (2)$$

$$\overline{a} = n_t^j + d_t^j, \qquad (3)$$

with l_t^j indicating the labor supply, d_t^j the time spent in day care centers and $\overline{a} \leq 1$ indicating the care time required by each child. Hereafter we assume for the sake of exposition that $\overline{a} = 1$. We posit that parents buy non-parental time and bear a resource cost per hour equal to $p(e_t)$, which is increasing in the quality level e_t enforced by the government, i.e. $p'(e_t) > 0$.

[4] In a background version of the paper (Casarico et al. 2011) we relax the assumption of perfect correlation between the two types of skills and consider a four-type model.

2.2 Output and evolution of skills' distribution

Output Q_t is produced according to the following linear technology:

$$Q_t = f_t^1 l_t^1 H^1 + f_t^2 l_t^2 H^2, \tag{4}$$

where f^j is the fraction of people of type j. Total population is normalized to 1 and the population growth rate is equal to 0. Output can be used for consumption or for day care services.

The dynamics of the fraction of high-skilled agents is described by:

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

The fraction of low skilled is then residually determined as $f_{t+1}^1 = 1 - f_{t+1}^2$.

3 Optimal public policy under a mixed tax system

In this section we characterize the optimal public policy chosen by a benevolent government. The public policy consists of the design of an optimal tax system and of the enforcement of an optimally chosen quality level of day care services.

In the tradition of the optimal income tax literature à la Mirrlees (1971), we assume that an individual's type and labor supply are not publicly observable but that labor income $Y_t^j = wH^j l_t^j$ is, where w denotes the wage in efficiency units, which is constant following the linear technology in equation (4). The non-observability assumption rules out the possibility for the government to levy type-specific lump-sum taxes but allows labor income to be taxed via a general (nonlinear) tax schedule $T\left(Y_t^j\right)$. We also assume that the government can subject the purchases of goods/services to a set of differentiated linear taxes. Thus, we focus on what is known in the literature as a "mixed" tax system. Choosing the consumption good c as the untaxed numeraire, the choice of the indirect tax structure boils down to the choice of the optimal tax/subsidy rate on day care services, which will be denoted by τ_t. Thus, the consumer price of one unit of day care services at time t is given by $p(e_t) + \tau_t$.

The government chooses the optimal fiscal policy maximizing the following objective function:

$$W=\sum_{t=0}^{\infty}\rho^{t}\sum_{j=1}^{2}f_{t}^{j}\cdot G\left(u\left(c_{t}^{j},z_{t}^{j},n_{t}^{j}\right)+\left(\pi^{j}(n_{t}^{j})\eta(H^{2})+(1-\pi^{j}(n_{t}^{j}))\eta(H^{1})\right)\right), \qquad (6)$$

where ρ is the government's discount factor, $G(\cdot)$ is a concave function which controls the degree of social aversion to inequality.

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. These constraints require that each agent must prefer the point on the income tax schedule intended for his type, rather than to misrepresent his true ability type and choose a point intended for another type. An agent misrepresenting his ability type is called mimicker. Following the bulk of the literature, we focus on the so-called "normal" case in which 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, from (3), $d_t^j = 1 - n_t^j$):

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

The government's problem can then be summarized by the following Lagrangian:

$$\begin{aligned}\pounds &= \sum_{t=0}^{\infty}\rho^t\sum_{j=1}^{2}f_t^j G\left(V_t^j\right)+\sum_{t=0}^{\infty}\rho^t\lambda_t\left(V_t^2-\widehat{V}_t^2\right)+\\ &\quad \sum_{t=0}^{\infty}\rho^t\mu_t\sum_{j=1}^{2}f_t^j\left(Y_t^j-B_t^j+\tau_t d_t^j\right)-\sum_{t=0}^{\infty}\rho^t\upsilon_t\left[f_{t+1}^2-\sum_{j=1}^{2}f_t^j\pi^j\left(n_t^j\right)\right],\end{aligned} \qquad (7)$$

where λ_t, μ_t and υ_t are a set of positive Lagrange multipliers associated respectively with the self-selection constraint, the government's budget constraint and the evolution of the skills' distribution.[5]

Below, we begin by characterizing the optimal tax structure and then move to the analysis of the optimal quality of day care services.

3.1 The optimal tax structure

As a measure of the distortions imposed by an optimal tax structure on the agents' labor supply we consider the concept of marginal effective tax rate, which is a measure of the distortion on labor supply produced by the combined effect of income and commodity taxation.

The marginal effective tax rate ($METR$) 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 j ($j = 1, 2$) is defined as:

$$METR_t^j \equiv T'\left(Y_t^j\right) + \left[\frac{\partial d_t^j}{\partial Y_t^j} + \left(1 - T'\left(Y_t^j\right)\right)\frac{\partial d_t^j}{\partial B_t^j}\right]\tau_t, \tag{8}$$

where $T'\left(Y_t^j\right)$ denotes the marginal income tax rate.

We can now state the following result:

Proposition 1 *Under a mixed tax system the optimal marginal effective tax rates can be expressed as:*

$$METR_t^1 = \frac{\lambda_t}{\mu_t f_t^1}\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} - \frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\right) - \frac{\upsilon_t}{\mu_t}\frac{\partial \pi^1}{\partial n_t^1}\left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0}, \tag{9}$$

$$METR_t^2 = -\frac{\upsilon_t}{\mu_t}\frac{\partial \pi^2}{\partial n_t^2}\left(\frac{dn_t^2}{dY_t^2}\right)_{dV_t^2=0}, \tag{10}$$

where $\left(\frac{dn_t^j}{dY_t^j}\right)_{dV_t^j=0} \equiv \frac{\partial n_t^j}{\partial Y_t^j} - \left(\frac{\partial V_t^j}{\partial Y_t^j}/\frac{\partial V_t^j}{\partial B_t^j}\right)\frac{\partial n_t^j}{\partial B_t^j}$.

Proof. See online Appendix. ■

Equation (9) provides an expression for the METR faced by low-skilled agents at an optimum. The first term on the right hand side reflects the distortion that the tax system ought to

[5] We assume that the government's budget is balanced year by year without recurring to debt.

impose on the labor supply of the low-skilled agents in order to prevent the high-skilled agents from becoming mimickers by choosing the (Y, B)-bundle intended for the low skilled. The sign of this self-selection term coincides with the sign of the expression within brackets. In standard models of nonlinear redistributive income taxation,[6] an 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 (9) is positive, therefore calling for a downward distortion on the labor supply of low-skilled agents.[7]

The second term on the right hand side of equation (9) has the same structure of that appearing on the right hand side of equation (10). Both correct for the intergenerational externality in human capital accumulation due to warm-glow altruism. 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 to 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 $t+1$. The sign of $\partial\pi^j/\partial n_t^j$ in equations (9) and (10) could be either positive or negative. In this discussion, we focus on the case in which $\partial\pi^1/\partial n_t^1 < 0$ and $\partial\pi^2/\partial n_t^2 > 0$. In this instance, at the margin and in the neighborhood of a social optimum, substituting time spent with children for time spent in day care centers increases (decreases) the probability that children of low(high)-skilled parents become high-skilled adults.[8] Under the plausible assumption that $\left(dn_t^j/dY_t^j\right)_{dV_t^j=0} < 0$ (i.e. an increase in labor supply is accomodated by adjusting downwards both uses of leisure time), the second term on the right hand side of equation (9) is negative and it tends to reduce the METR faced by low-skilled agents, whereas the METR in equation (10) faced by high-skilled agents is positive. These corrections represent, on the one

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

[7] Note, 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 condition for agent-monotonicity. In our setting, this is not enough since a high-skilled mimicker and a true low-skilled agent do not only differ with respect to their labor supply but also with respect to the amount available for private consumption (once expenses on day care services have been subtracted). A more thorough discussion of this issue is provided in a background version of this paper (see Appendix A in Casarico et al. 2011).

[8] We recall 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.

side, a way to induce low-skilled agents to work more and substitute consumption for leisure time (including that spent with children); on the other side, an incentive to high-skilled agents to under-provide labor supply in order to make them spend more time with their children.

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

$$\tau_t = \frac{\frac{\lambda_t}{\mu_t}\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(d_t^1 - \widehat{d}_t^2\right)}{\sum_{j=1}^{2}\frac{\partial \widetilde{d}_t^j}{\partial \tau_t}f_t^j} + \frac{\upsilon_t}{\mu_t}\sum_{j=1}^{2}\frac{\partial \pi^j}{\partial n_t^j}\zeta_t^j, \tag{11}$$

where $\zeta_t^j \equiv \frac{\partial \widetilde{d}_t^j}{\partial \tau_t}f_t^j / \sum_{j=1}^{2}\frac{\partial \widetilde{d}_t^j}{\partial \tau_t}f_t^j$.

Proof. See online Appendix. ■

The denominator of the first term on the right hand side of (11) is negative and it provides a measure of the deadweight loss generated by distortionary commodity taxation. The numerator depends on the difference between the amount of day care used by a true low skilled and by a high-skilled mimicker. Since a mimicker's labor supply is lower than that of a true low skilled, it is reasonable to assume that $d_t^1 - \widehat{d}_t^2 > 0$. Thus, the first term on the right hand side of (11) calls for a subsidy on the purchase of day care services. Intuitively, given that $d_t^1 > \widehat{d}_t^2$, 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 and 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).

As to the second term of equation (11), ζ_t^j represents the normalized change, generated by a marginal increase in τ_t, in the compensated demand for day care services by agents of skill type j. This second term is reminiscent of a similar term appearing in (9) and (10). The main difference is that in (11) we take a sum over $j = 1, 2$. This is due to the different degree of sophistication of the available tax instruments. Since labor income is assumed to be taxable nonlinearly, 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. Thus, a single tax rate, τ_t, has to be tailored in a way that strikes a balance between the adjustments ideally required to correct the behavior of the low-skilled and of the high-skilled agents. Since $\zeta_t^j > 0$ for all j, whereas 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 opposite for high- and low-skilled agents. Thus, the optimal value of τ_t tends to be pushed up (down) by the concern to affect the time allocation of high(low)-skilled parents. Note also that a high value of ζ_t^j reflects an instance in which the commodity tax is a very effective instrument to alter the demand for day care services by agents of type j. It is therefore very effective also in influencing the amount of time they spend with children. In this case, the optimal value chosen for τ_t will tend to reflect more strongly how it can be used to indirectly affect, in the socially optimal direction, the time spent with children by parents of skill type j.

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

3.2 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 for an agent of type j at time t (keeping the consumer price of day care services fixed), we have:

$$MRS_{ec}^{j,t} \equiv \frac{\partial V_t^j}{\partial e_t} / \frac{\partial V_t^j}{\partial B_t^j} = \left[\eta\left(H^2\right) - \eta\left(H^1\right)\right] \frac{\partial \pi^j}{\partial e_t} / \frac{\partial V_t^j}{\partial B_t^j}. \quad (12)$$

With the help of (12), Proposition 3 provides a characterization of the optimal quality of day care services under a mixed tax system.

Proposition 3 *Under a mixed tax system the optimal quality of day care services abides by the following rule:*

$$\begin{aligned}\sum_{j=1}^{2} f_t^j MRS_{ec}^{j,t} =& p'\left(e_t\right) \sum_{j=1}^{2} d_t^j f_t^j - \frac{\upsilon_t}{\mu_t} \sum_{j=1}^{2} \left(\frac{d\pi^j}{de_t}\right)_{dV^j=0} f_t^j + \\ & \frac{\lambda_t}{\mu_t} \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \left(\widehat{MRS}_{ec}^{2,t} - MRS_{ec}^{1,t}\right) - \tau_t \sum_{j=1}^{2} \left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0} f_t^j,\end{aligned} \quad (13)$$

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:*

$$\left(\frac{d\pi^j}{de_t}\right)_{dV^j=0} \equiv \frac{\partial \pi^j}{\partial e_t} + \left(\frac{\partial n_t^j}{\partial e_t} - \frac{\partial n_t^j}{\partial B_t^j}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}\right)\frac{\partial \pi^j}{\partial n_t^j}, \tag{14}$$

$$\left(\frac{\partial d_t^j}{\partial e_t}\right)_{dV^j=0} \equiv \frac{\partial d_t^j}{\partial e_t} - \frac{\partial d_t^j}{\partial B_t^j}\frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}}. \tag{15}$$

Proof. See online Appendix. ■

Eq. (13) 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 (13) 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 (13) represents the additional resource cost of raising the quality of day care services, for given number of day care hours. Discounting 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 (13) describe how an optimizing policy maker should deviate from the first-best rule to take into account self-selection problems and externalities. One can note that the presence of the last term on the right side of (13) does not challenge this interpretation because, as evident from (11), a commodity tax/subsidy on day care services can only be justified by self-selection reasons or externalities.

The second term on the right side of (13) reflects how a compensated variation in the quality of day care services can be used for externality-correction purposes. 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 originates from 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 $(d\pi^j/de_t)_{dV^j=0}$, which also reflects how parents vary the time spent with their children following a reduction in disposable income intended to leave their utility unchanged. The sign of $(d\pi^j/de_t)_{dV^j=0}$ is therefore in general ambiguous. However, making the assumption that the direct effect of an increase in the quality level dominates the indirect ones, $(d\pi^j/de_t)_{dV^j=0} > 0$ and the sign of the second term on the right side of (13) is negative, calling for an increase in the second-best efficient level of day care quality.

As to the third term on the right side of (13), it 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 of a true low skilled. Assuming that, having more time to devote to non-market activities, a mimicker spends more time with his child and therefore less money on day care services, the marginal utility of consumption is likely to be lower for a mimicker than for a true low skilled. Taking this into account, (12) tends to imply that the marginal willingness to pay for increased quality is larger for a mimicker than for a true low skilled.[9] According to the third term on the right hand side of equation (13), this determines 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 therefore tighten the self-selection constraint.

Finally, the last term on the right side of (13) 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 (13) 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.

4 Numerical analysis

Since the formulas derived in the previous Section have in most cases components working in opposite directions, here we develop a numerical analysis. Its purpose is to illustrate the quantitative relevance of including child care in the human capital production function, both

[9] It is however clear from (12) that one should also consider how the numerator of the expression defining the marginal willingness to pay for quality differs for a mimicker and for a true low skilled. For simplicity, in our discussion here we disregard the possibility that this effect more than offsets the effect that works through the difference in the denominators.

in terms of the optimal values of the policy variables and in terms of the welfare loss caused by setting policy neglecting its effect on child care arrangements and on the skill formation process. We maintain the two-period, two-type structure developed in the theoretical part. We focus on the steady state and therefore we omit time subscripts.

4.1 Functional forms

To perform the numerical exercise we need to 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.

As to the utility function, we assume:

$$U^j = \alpha c^j + \beta \log z^j + \gamma \log n^j + \delta \left[\pi^j(n^j) \log H^2 + (1 - \pi^j(n^j)) \log H^1 \right], \qquad (16)$$

with $\alpha + \beta + \gamma + \delta = 1$, $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\delta > 0$.[10]

The probability of becoming high skilled is:

$$\pi^j(n^j) = \sigma^j \frac{x^j}{1 + x^j}, \qquad (17)$$

where $\sigma^j > 0$ is a type-specific parameter capturing in a reduced form other factors besides the quality of early childhood enviroment x^j, which can affect the probability of becoming a high-skilled agent. The assumption is that, for a given value of x^j, high skilled parents can positively influence the chances of their offspring to become high skilled, for instance, through the transmission of genetic ability, higher income, and better schooling.

Quality x^j is produced combining parental time and day care services using the following constant elasticity of substitution function:

$$x^j = \left[\left(n^j H^j \right)^\nu + \left(d^j e \right)^\nu \right]^{\frac{1}{\nu}}. \qquad (18)$$

Note that equations (17) and (18), jointly with (3), imply that the relationship between π^j and n^j is hump shaped. When parental time n^j is low, its increase (i.e. a decrease in time

[10] This objective function is of the Greenwood-Hercowitz-Huffman type. Note that it is linear in consumption, which implies that there is no income effect on the labor supply, and it is often used in the optimal taxation literature (e.g. see Kleven et al. (2009)).

spent in day care) generates a rise in the probability that the children will be high skilled in the next period. When it is high, a further increase determines a decline in the same probability.

The relationship between the price p of day care services (i.e. the price without tax or subsidy components) and their quality e is specified as follows

$$p = \left(\frac{e}{\omega}\right), \tag{19}$$

where $\omega > 0$ is a scale parameter. The higher is the quality of day care services, the higher is the price.

As to the government, the objective function is the following:

$$W = \sum_{t=0}^{\infty} \rho^t \sum_{j=1}^{2} f_t^j \cdot \frac{(U^j)^\psi}{\psi}. \tag{20}$$

The parameter $\psi \in (-\infty, 1)$ measures the degree of social aversion to inequality: the lower is ψ, the higher is the aversion to inequality.[11]

4.2 Parameterization and Calibration

We interpret each period in the model as having a length of 25 years. We set $\bar{a} = 0.24$, which means that over a period of 25 years, 6 years are spent receiving child care. Though we only aim at performing a numerical example, we try to make it realistic and we rely, whenever possible, on actual data. More specifically, we use Italian data.[12] We classify as agents of type 1 all the individuals with no college education (up to ISCED Level 4 included), whereas agents of type 2 are those with college education and above (ISCED Level 5 and above). We normalize to 1 the market productivity of agents of type 1: $H^1 = 1$. In order to set H^2, we use IT-Silc 2004 data and calculate hourly gross wages for the two groups. As the resulting ratio is equal to 1.5, we set $H^2 = 1.5$.

Utility function

Concerning the parameters of the utility function, we select them in order to match data on time allocation. Assuming, as it is usually done (e.g. Juster 1985), that non-personal

[11] Given that we focus on the steady state, the discount factor ρ is just a scale parameter which does not affect the optimal policies.

[12] Data used are not all available for a single year. The reference period is $1998 - 2004$.

time available for discretionary use amounts to 100 hours per week, according to HETUS (Harmonised Time Use Survey) data for Italy for 2002-2003 for the age group 25-50 (the parents' group in the model), we have: $l^1 = 31\% < l^2 = 36\%$, $n^1 = n^2 = 4\%$, $z^1 = 65\% > z^2 = 60\%$.[13] It should be stressed that high-skilled agents work more even though they do not devote less time to their children (see also Guryan et al. 2008). Time use data provide six targets (l^1, l^2, n^1, n^2, z^1, z^2). Two of them, e.g. leisure time z^1 and z^2, can be residually determined through the time constraint (2). To match the other four targets we only have three parameters: α, β and γ, given that $\alpha + \beta + \gamma + \delta = 1$ and δ is residually determined by this constraint. We set these parameters in order to match time devoted to children by type 1 and type 2 agents (n^1 and n^2, respectively) and the labor supply of type 1 agents l^1.[14] The implied value of the labor supply of type 2 agents l^2 is equal to 33%. This value is slightly lower than the value we observe in the data but it qualitatively matches well the behavior of the labor supply across educational groups.

Child care and human capital

The parameter ν in equation (18) determines the elasticity of substitution between parental time n^j and day care services d^j in the production of the quality of the early childhood environment x^j. This elasticity is equal to $1/(1-\nu)$. When $\nu = 1$ ($\nu = -\infty$), n^j and d^j are perfect substitutes (perfect complements). To the best of our knowledge, 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. As a benchmark we consider the case $\nu = 0.98$ and in the online Appendix we provide a sensitivity analysis.

The parameters of the probability function (17) σ^j are chosen to match the probabilities of becoming a high-skilled agent, namely $\pi^1 = 0.065$ and $\pi^2 = 0.59$,[15] which, together with the assumed values for time devoted to child care n^1 and n^2, imply fractions of low and

[13] A few remarks are important in interpreting these data. First, we recall that in our model parents devote time to the child only during his first six years of life. In collecting time use data we therefore only consider time devoted to children over this time span, whereas the overall time endowment refers to a 25-year period. This partly explains the low share of time devoted to child care. Moreover, parental time with children is defined as the sum of the minutes devoted to primary and secondary child care: this amount of time is lower than the total time spent with children but it better captures deliberate child care by parents. Finally, leisure is 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.

[14] The resulting values of α, β and γ are, respectively, 0.78, 0.19, 0.01.

[15] See Giuliano (2008). Similar estimates are provided by Checchi et al. (1999) on older data (1985 rather than 1998). The calibration delivers $\sigma^1 = 0.31$ and $\sigma^2 = 2.69$.

high-skilled agents respectively equal to 0.86 and 0.14.[16]

As to the quality of day care e, we recall that in our model there is perfect correlation between productivity on the market and ability to rear children. Accordingly, a plausible empirical counterpart of e can be the productivity of people providing non-parental care, who typically in Italy have high-school education. Using IT-Silc data for 2004, we calculate the hourly gross wages for workers with a high-school diploma. The ratio between the wage for this group and the wage for the group of agents of type 1 (which also include agents without a high school diploma) is 1.09 and therefore we set $e = 1.09$.

We now turn to the relationship between the quality of day care services e and the price p. The scale parameter ω is chosen so that the ratio between the price of day care services and the wage of agents of type 1 $p/(wH^1)$ is equal to 31%.[17]

Fiscal variables

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

$$T^j = \chi \cdot (Y^j)^{1+\phi}, \tag{21}$$

with $Y^j = wH^j l^j$. The parameter χ is chosen to match an average tax rate equal to 38%. The parameter ϕ is set equal to 0.5 to match a ratio between the marginal and the average tax rate equal to 1.5.[18] The current tax system is also characterized by an *ad valorem* tax on day care, whose rate τ^d is set equal to -87% (see Zollino (2008)), and by an *ad valorem* tax on consumption, whose rate τ^c is equal to 22% (see McDaniel 2007 tax data series for 2004). We allow the government to also use a lump-sum transfer, which is residually determined in order to have a balanced government budget.

Note that the (parametric) fiscal system described above is simply used for the purpose of calibrating the model: the social planner will choose a fully nonlinear income tax to maximize

[16] According to the OECD (2011), the share of Italian graduates in the age group 25-64 was 15% in 2009.

[17] The unitary cost of day care services is calculated averaging OECD data on the Italian public and private per-child expenditure on pre-primary education (children aged 3-6) and Bank of Italy data presented in Zollino (2008) on the public and private per-child expenditure on day care (children aged 0-3). The obtained per child cost is then converted into an hourly cost assuming that care is available for 11 months, 8 hours per day. The hourly cost is finally divided by the wage rate of type 1 agents: the resulting value is 31%. The calibration delivers $\omega = 0.28$.

[18] OECD data for a two-earner married couple, with one spouse earning the average wage, the other spouse earning 33% of the average wage, and two children.

the social welfare function (20), in which the social aversion to inequality ψ is set to -2.[19] Moreover the tax rate on consumption is normalized to zero.

4.3 Simulation: optimal policy

In order to analyze how the presence of child care in the human capital production function affects the design of the optimal public policy, we introduce as a term of comparison a standard model in which day care arrangements do not have an impact on human capital accumulation, that is $\pi^j(n_t^j) = \pi^j$. The calibration procedure is analogous to the one implemented in the previous section.[20]

Table 1: Optimal policies

	Standard	Quality fixed	Quality optimized
e	1.09	1.09	1.50
T^1/Y^1	-0.99%	-1.28%	-0.56%
T^2/Y^2	18.32%	18.15%	19.40%
T'^1	7.38%	7.06%	11.91%
T'^2	0.52%	3.16%	1.67%
$METR^1$	6.44%	6.23%	9.97%
$METR^2$	0.00%	2.69%	0.61%
τ^d	-36.28%	-31.01%	-53.41%

We compute the optimal policies and we present the results in Table 1. In the first column we report the results of the simulation performed in a standard model in which the process of skills' transmission is entirely exogenous. In the second and third column we report the optimal policies for the model in which the skills' transmission depends on parental time.

[19] We perform a sensitivity analysis on ψ in the online Appendix.

[20] The only difference concerns the parameters of the utility function. Since the probability of becoming high skilled π^j does not depend on parental time n^j, the parameter δ does not affect the allocation of time, and we normalize it to 0. Thus, to match the allocation of time of the two groups we only have two parameters α and β, given the restriction $\alpha + \beta + \gamma = 1$. These two parameters are used to match the labor supply and the time devoted to children by the low skilled l^1 and n^1; l^2 and n^2 are generated by the model.

More precisely, in the second column we assume that the quality of day care is fixed at a level equal to the productivity of people currently working in the day care sector, i.e. $e = 1.09$; in the third column we include the quality of day care among the government's policy instruments. The comparison of the optimal policies across the different cases allow us to isolate the roles played by the relationship between day care arrangements and the distribution of skills, and by the quality of day care, which are the novelties of our model.

We first highlight that the main difference in results between the standard model and ours is the degree of subsidization of day care. Also the marginal income tax rates change, though at a smaller extent. This shows that it is the direct instrument of taxing/subsidizing the purchases of day care services which is mostly affected by the intergenerational externality in the skill transmission.

We first compare the results in column 1 and 2, starting from the marginal effective tax rates. To this end we recall equations (9) and (10) and note that they hold also in the standard model, in which, however, $\frac{\partial \pi^j}{\partial n^j} = 0$. The marginal effective tax rate on the high-skilled individuals ($METR^2$) is zero in the standard model, as it is customary, whereas it is positive in our model. This positive marginal effective tax rate corrects for the externality stemming from the intergenerational transmission of skills: high-skilled agents are induced to spend more time with their children as this additional time has - at the optimum - a positive impact on the probability that they will be high skilled. The marginal effective tax rate on the low- skilled individuals ($METR^1$) is slightly lower in our model compared to the standard one, given that at the optimum the time they spend with their children generates a negative externality on the probability of becoming high skilled.

As to the tax on day care τ^d, it is negative in both column 1 and 2, calling for a subsidy on day care. This subsidy is lower in our model than in the standard one. The intuition can be grasped looking at equation (11).[21] The denominator of this equation is negative, given that an increase in taxation on day care reduces its demand. Moreover, a low-skilled agent buys a higher amount of day care than a high-skilled mimicker and therefore self-selection considerations always call for a negative value of τ^d. In the standard model there are no additional determinants of τ^d besides self-selection. In our model the setting of the tax rate

[21] We here consider an ad valorem tax τ^d, whereas equation (11) refers to an excise tax τ. Note that setting $\tau^d = \tau/p$, we obtain for an ad valorem tax an equation which is equivalent to (11).

on day care depends also on how it affects the probability that children of both types become high-skilled individuals. In our simulation the need to discourage the high skilled from using day care services prevails on the need to decrease parental time provided to children by low skilled parents and this implies that τ^d is less negative in our model that in the standard one.

Column 3 of Table 1 shows the effects of including the quality of day care services e among the policy instruments of the government. The range of values of e goes from $H^1 = 1$, which is the productivity of low-skilled agents, to $H^2 = 1.5$, which is the productivity of high-skilled agents. We find that the government sets the optimal quality of day care services at the highest possible value, that is H^2. It follows that $METR^2$ is lower than the one observed in column 2 which reports the value for the case in which the quality of day care services is exogenously fixed at a lower level. When the quality of day care is higher, a reduction in parental time by high-skilled agents determines a smaller decrease in the probability that their children will become high skilled. $METR^1$ is instead higher when the quality is endogenous. This result occurs since, as we pointed out in Section 3.2, the marginal willingness to pay for increased quality is larger for a mimicker than for a true low skilled. Therefore, an increase in day care quality tends to tighten the binding self-selection constraint which prevents high-skilled agents to mimic low-skilled ones. To counteract this effect the government raises the distortion, i.e. the marginal effective tax rate faced by the low-skilled agents. Finally, as to the tax on day care τ^d, we observe that the size of the subsidy increases significantly, since its positive impact on low-skilled agents weighs much more than the small negative impact on the high skilled in determining the overall effect.

4.4 Simulation: welfare effects

In this Section we assess the welfare implications of choosing the optimal public policy ignoring child care arrangements as an input of the human capital production function.

To this end we use the following two-step procedure. First, we implement in our model (in which the probability of becoming high skilled $\pi^j(n_t^j)$ is given by equation (17)) the optimal policies calculated in the standard model (in which $\pi^j(n_t^j) = \pi^j$). This amounts to describing the government as using a *mistaken* environment when selecting the optimal public policies. Taking these policies as given, agents maximize their objective function in the *right* environment, in which their choices affect the human capital of the next generation. We find that the incentive compatibility constraint of high-skilled agents is violated: the high skilled find it optimal to mimic the low skilled and the mimicking generates a deficit. We calculate the welfare level in this case. Second, we compute the deficit which guarantees this very same level of welfare, when however the government uses the right environment, both allowing and not allowing for an endogenous choice of the quality of day care. The difference in the deficit generated by the government which uses the right environment and by the government which uses the wrong environment represents a revenue-based measure of the welfare loss: the loss amounts respectively to 3.8% of GDP when the quality of day care is fixed and to 4.3% of GDP when the quality of day care is optimally chosen.[22] This suggests that ignoring the role of child care arrangements in the process of human capital accumulation when setting the optimal public policies can generate a sizable welfare loss.

5 Conclusions

This paper has characterized the optimal structure of a mixed tax system and the optimal level of quality of day care in an OLG model in which parental choices over child care arrangements affect the probability that a child becomes a high-skilled adult in a type-specific way. As far as we know, this is the first model to include child care arrangements in the human capital production function to analyze optimal public policies.

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. The Pigouvian terms work in the direction of lowering (increasing) the marginal *effective* tax rate when, at the optimum, parental time devoted to children generates a negative (positive) externality on the probability of becoming high skilled. As to the tax on day care services, given that it is assumed to be linear, it has to be tailored in a way that strikes a balance between the adjustments ideally required to correct the behavior of the different types of agents. As far as the optimal choice of the quality of day care is concerned, this represents a new policy instrument motivated by the existence of a link between child

[22] GDP is measured at the optimum of the government's problem when the right environment is used.

care arrangements and human capital accumulation. We find that the optimal quality is determined by equating the total private marginal benefits of a quality increase to its marginal cost, adjusted for the presence of additional terms capturing respectively the budgetary impact of a change in demand for day care services, the intergenerational externality in human capital accumulation, and the self-selection constraint.

We also perform a numerical analysis to illustrate the working of the model, the ensuing optimal public policies and the welfare loss from designing the public policy without taking into account the effects of parental time on children's human capital. We find that the presence of child care arrangements in the human capital production function mainly influences the degree of subsidization of day care. Designing the public policy without taking into account the effects of parental time on children's human capital causes a welfare loss ranging from 3.8% to 4.3% of GDP depending on whether the quality is optimally set or not. The numerical analysis is just meant to give a first assessment of the relevance of introducing child care arrangements in the human capital production function. An interesting extension would concern the use of a multiperiod OLG model, with many types and imperfect correlation between the ability to rear children and the market ability of parents. This setting would allow to draw more precise quantitative predictions on the optimal policies, but would require a deeper knowledge of the technology of skill formation than the one currently available. Although there is an increasing literature on the effects of parental time on children's human capital, to the best of our knowledge, there is no consensus on the effects of day care arrangements on children's skills conditional on parental education, and on the degree of correlation between the ability to rear children and the market ability of parents. The analysis presented in this paper, despite its simplifying assumptions, shows that the existence of a link between parental time devoted to children and children human capital can be relevant for the design of public policies and provides the ground and motivation for further empirical analysis to improve our knowledge of how alternative child care arrangements can shape the human capital accumulation process.

References

Balestrino, A., 2000. Mixed tax systems and the public provision of private goods. Int. Tax Public Finance 7, 463–478.

Balestrino, A., Cigno, A., Pettini, A., 2002. Endogenous fertility and the design of family taxation. Int. Tax Public Finance 9, 175–193.

Balestrino, A., Cigno, A., Pettini, A., 2003. Doing wonders with an egg: Optimal redistribution when households differ in market and non-market abilities. J. Public Econ. Theory 5, 479–498.

Bernal, R., 2008. The effect of maternal employment and child care on children's cognitive development. Int. Econ. Rev. 49, 1173–1209.

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

Bernal, R., Keane, M., 2011. Child care choices and children's cognitive achievement: the case of single mothers. J. Lab. Econ. 29, 459–512.

Blomquist, S., Christiansen, V., Micheletto, L., 2010. Public provision of private goods and nondistortionary marginal tax rates. Am. Econ. J.: Ec: Economic Policy 2, 1–27.

Casarico, A., Micheletto, L., Sommacal, A., 2011. Intergenerational Transmission of Skills during Childhood and Optimal Public Policy. CESifo Working Paper Series 3343. CESifo Group Munich.

Casarico, A., Sommacal, A., 2012. Labor income taxation, human capital and growth: the role of child care. Scand. J. Econ. 114, 1182–1207.

Checchi, D., Ichino, A., Rustichini, A., 1999. More equal but less mobile? education financing and intergenerational mobility in italy and in the us. J. Public Econ. 74, 351–393.

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

Cunha, F., Heckman, J., 2007. The technology of skill formation. Amer. Econ. Rev. 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.

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

Giuliano, P., 2008. Culture and the family: An application to educational choices in Italy. Rivista Politica Econ. 98, 3–38.

Glomm, G., Kaganovich, M., 2008. Social security, public education and the growth inequality relationship. Europ. Econ. Rev. 52, 1009–1034.

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

Guryan, J., Hurst, E., Kearney, M., 2008. Parental education and parental time with children. J. Econ. Perspect. 22, 23–46.

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. Working Papers 13016. NBER.

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.

Kleven, H.J., Kreiner, C.T., Saez, E., 2009. The optimal income taxation of couples. Econometrica 77, 537–560.

Kopczuk, W., 2013. Incentive effects of inheritances and optimal estate taxation. Amer. Econ. Rev..: Pap. and Proc. 103, 472–477.

Li, W., Sarte, P.D., 2004. Progressive taxation and long-run growth. Amer. Econ. Rev. 94, 1705–1716.

McDaniel, C., 2007. Average tax rates on consumption, investment, labor and capital in the oecd 1950-2003. Mimeo.

Mirrlees, J.A., 1971. An exploration in the theory of optimum income taxation. Rev. Econ. Stud. 38, 175–208.

OECD, 2011. Education at a glance 2011.

Ruhm, C., 2004. Parental employment and child cognitive development. J. Human Res. 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. J. Public Econ. 17, 213–240.

Zollino, F., 2008. Il difficile accesso ai servizi di istruzione per la prima infanzia in Italia: i fattori di offerta e di domanda. Questioni di Economia e Finanza (Occasional Papers) 30. Bank of Italy, Economic Research Department.

A Proofs

A.1 Proof of Proposition 1

Agents choose their labor supply maximizing $V\left(Y_t^j, B_t^j; wH^j\right)$ subject to the link between pre-tax earnings and post-tax earnings available for goods expenditure implied by the direct tax schedule. Accordingly, the condition $\frac{\partial V_t^j}{\partial B_t^j}\left(1 - T'\left(Y_t^j\right)\right) + \frac{\partial V_t^j}{\partial Y_t^j} = 0$ is satisfied at an optimum. This allows us to define the (implicit) marginal income tax rate faced by an agent as:

$$T'\left(Y_t^j\right) = 1 + \left(\frac{\partial V_t^j}{\partial Y_t^j} / \frac{\partial V_t^j}{\partial B_t^j}\right). \tag{A.1}$$

Thus, we can rewrite equation (8) as:

$$METR_t^j \equiv T'\left(Y_t^j\right) + \left(\frac{dd_t^j}{dY_t^j}\right)_{dV_t^j=0} \tau_t, \tag{A.2}$$

where, since $d_t^j = 1 - n_t^j$, $\left(\frac{dd_t^j}{dY_t^j}\right)_{dV_t^j=0} = -\left(\frac{dn_t^j}{dY_t^j}\right)_{dV_t^j=0}$

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

$$f_t^1 G'\left(V_t^1\right)\frac{\partial V_t^1}{\partial Y_t^1} = \lambda_t \frac{\partial \widehat{V}_t^2}{\partial Y_t^1} + \left[-\upsilon_t \frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial Y_t^1} - \mu_t\left(1 + \tau_t \frac{\partial d_t^1}{\partial Y_t^1}\right)\right] f_t^1; \tag{A.3}$$

$$f_t^1 G'\left(V_t^1\right)\frac{\partial V_t^1}{\partial B_t^1} = \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1} + \left[-\upsilon_t \frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1} - \mu_t\left(-1 + \tau_t \frac{\partial d_t^1}{\partial B_t^1}\right)\right] f_t^1. \tag{A.4}$$

Dividing (A.3) by (A.4) and multiplying the result by the right hand side of (A.4), we get:

$$\begin{aligned} &\frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\left\{\lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1} + \left[-\upsilon_t \frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1} - \mu_t\left(-1 + \tau_t \frac{\partial d_t^1}{\partial B_t^1}\right)\right] f_t^1\right\} \\ &= \lambda_t \frac{\partial \widehat{V}_t^2}{\partial Y_t^1} + \left[-\upsilon_t \frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial Y_t^1} - \mu_t\left(1 + \tau_t \frac{\partial d_t^1}{\partial Y_t^1}\right)\right] f_t^1. \end{aligned} \tag{A.5}$$

Using (A.1) to collect terms in (A.5) gives:

$$T'\left(Y_t^1\right) = -\left(\frac{dd_t^1}{dY_t^1}\right)_{dV_t^1=0} \tau_t + \frac{\lambda_t}{\mu_t f_t^1}\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(\frac{\frac{\partial \widehat{V}_t^2}{\partial Y_t^1}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} - \frac{\frac{\partial V_t^1}{\partial Y_t^1}}{\frac{\partial V_t^1}{\partial B_t^1}}\right) + -\frac{\upsilon_t}{\mu_t}\frac{\partial \pi^1}{\partial n_t^1}\left(\frac{dn_t^1}{dY_t^1}\right)_{dV_t^1=0}. \tag{A.6}$$

Using (A.6) to substitute terms in (A.2) gives (**??**).

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

$$\left[f_t^2 G'\left(V_t^2\right)+\lambda_t\right]\frac{\partial V_t^2}{\partial Y_t^2}=\left[-\upsilon_t\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial Y_t^2}-\mu_t\left(1+\tau_t\frac{\partial d_t^2}{\partial Y_t^2}\right)\right]f_t^2; \tag{A.7}$$

$$\left[f_t^2 G'\left(V_t^2\right)+\lambda_t\right]\frac{\partial V_t^2}{\partial B_t^2}=\left[-\upsilon_t\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right]f_t^2. \tag{A.8}$$

Dividing (A.7) by (A.8) and multiplying the result by the right hand side of (A.8), we get:

$$\frac{\frac{\partial V_t^2}{\partial Y_t^2}}{\frac{\partial V_t^2}{\partial B_t^2}}\left[-\upsilon_t\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t\frac{\partial d_t^2}{\partial B_t^2}\right)\right]=-\upsilon_t\frac{\partial \pi^2}{\partial n_t^2}\frac{\partial n_t^2}{\partial Y_t^2}-\mu_t\left(1+\tau_t\frac{\partial d_t^2}{\partial Y_t^2}\right). \tag{A.9}$$

Using (A.1) to collect terms in (A.9) gives:

$$T'\left(Y_t^2\right)=-\left(\frac{dd_t^2}{dY_t^2}\right)_{dV_t^2=0}\tau_t-\frac{\upsilon_t}{\mu_t}\frac{\partial \pi^2}{\partial n_t^2}\left(\frac{dn_t^2}{dY_t^2}\right)_{dV_t^2=0}. \tag{A.10}$$

We can then use (A.10) to substitute terms in (A.2) and in this way obtain (**??**).

A.2 Proof of Proposition 2

The first order condition for τ_t is given by:

$$\sum_{j=1}^{2}\frac{\partial V_t^j}{\partial \tau_t}f_t^j G'\left(V_t^j\right)\mu_t\sum_{j=1}^{2}\left(d_t^j+\tau_t\frac{\partial d_t^j}{\partial \tau_t}\right)f_t^j+\lambda_t\left(\frac{\partial V_t^2}{\partial \tau_t}-\frac{\partial \widehat{V}_t^2}{\partial \tau_t}\right)=-\upsilon_t\sum_{j=1}^{2}f_t^j\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial \tau_t}. \tag{A.11}$$

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:

$$-\sum_{j=1}^{2}d_t^j\frac{\partial V_t^j}{\partial B_t^j}f_t^j G'\left(V_t^j\right)+\mu_t\sum_{j=1}^{2}\left(d_t^j+\tau_t\frac{\partial d_t^j}{\partial \tau_t}\right)f_t^j+\lambda_t\left(-d_t^2\frac{\partial V_t^2}{\partial B_t^2}+\widehat{d}_t^2\frac{\partial \widehat{V}_t^2}{\partial B_t^1}\right)=-\upsilon_t\sum_{j=1}^{2}f_t^j\frac{\partial \pi^j}{\partial n_t^j}\frac{\partial n_t^j}{\partial \tau_t}. \tag{A.12}$$

Multiplying (A.4) and (A.8) by, respectively, d_t^1 and d_t^2, we can find the following two expressions for $-f_t^1 G'\left(V_t^1\right)d_t^1\frac{\partial V_t^1}{\partial B_t^1}$ and $-\left[f_t^2 G'\left(V_t^2\right)+\lambda_t\right]d_t^2\frac{\partial V_t^2}{\partial B_t^2}$:

$$-f_t^1 G'\left(V_t^1\right)d_t^1\frac{\partial V_t^1}{\partial B_t^1}=-\lambda_t d_t^1\frac{\partial \widehat{V}_t^2}{\partial B_t^1}-\left[-\upsilon_t\frac{\partial \pi^1}{\partial n_t^1}\frac{\partial n_t^1}{\partial B_t^1}-\mu_t\left(-1+\tau_t\frac{\partial d_t^1}{\partial B_t^1}\right)\right]f_t^1 d_t^1. \tag{A.13}$$

$$-\left[f_t^2 G'\left(V_t^2\right)+\lambda_t\right] d_t^2 \frac{\partial V_t^2}{\partial B_t^2}=-\left[-\upsilon_t \frac{\partial \pi^2}{\partial n_t^2} \frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t \frac{\partial d_t^2}{\partial B_t^2}\right)\right] f_t^2 d_t^2; \qquad (A.14)$$

Substituting (A.13) and (A.14) into (A.12) 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:

$$-\left[-\upsilon_t \frac{\partial \pi^2}{\partial n_t^2} \frac{\partial n_t^2}{\partial B_t^2}-\mu_t\left(-1+\tau_t \frac{\partial d_t^2}{\partial B_t^2}\right)\right] f_t^2 d_t^2-\left[-\upsilon_t \frac{\partial \pi^1}{\partial n_t^1} \frac{\partial n_t^1}{\partial B_t^1}-\mu_t\left(-1+\tau_t \frac{\partial d_t^1}{\partial B_t^1}\right)\right] f_t^1 d_t^1+$$
$$\mu_t \sum_{j=1}^2\left[d_t^j+\tau_t\left(\frac{\partial \widetilde{d}_t^j}{\partial \tau_t}-d_t^j \frac{\partial d_t^j}{\partial B_t^j}\right)\right] f_t^j-\lambda_t d_t^1 \frac{\partial \widehat{V}_t^2}{\partial B_t^1}+\lambda_t \widehat{d}_t^2 \frac{\partial \widehat{V}_t^2}{\partial B_t^1}=-\upsilon_t \sum_{j=1}^2 f_t^j \frac{\partial \pi^j}{\partial n_t^j} \frac{\partial n_t^j}{\partial \tau_t}, \qquad (A.15)$$

which can be simplified to obtain:

$$\sum_{j=1}^2 f_t^j \tau_t \frac{\partial \widetilde{d}_t^j}{\partial \tau_t}=-\frac{\upsilon_t}{\mu_t}\left[\sum_{j=1}^2 f_t^j \frac{\partial \pi^j}{\partial n_t^j} \frac{\partial n_t^j}{\partial \tau_t}+\frac{\partial \pi^2}{\partial n_t^2} \frac{\partial n_t^2}{\partial B_t^2} f_t^2 d_t^2+\frac{\partial \pi^1}{\partial n_t^1} \frac{\partial n_t^1}{\partial B_t^1} f_t^1 d_t^1\right]+\frac{\lambda_t}{\mu_t} \frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(d_t^1-\widehat{d}_t^2\right). \qquad (A.16)$$

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 (A.16) in a more compact form as:

$$\sum_{j=1}^2 f_t^j \tau_t \frac{\partial \widetilde{d}_t^j}{\partial \tau_t}=\frac{\lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1}}{\mu_t}\left(d_t^1-\widehat{d}_t^2\right)-\frac{\upsilon_t}{\mu_t} \sum_{j=1}^2 f_t^j \frac{\partial \pi^j}{\partial n_t^j} \frac{\partial \widetilde{n}_t^j}{\partial \tau_t}. \qquad (A.17)$$

Then, exploiting the time-constraint $d_t^j=1-n_t^j$, 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 equation:

$$\tau_t=\frac{\lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1}\left(d_t^1-\widehat{d}_t^2\right)-\upsilon_t \sum_{j=1}^2 \frac{\partial \pi^j}{\partial d_t^j} \frac{\partial \widetilde{d}_t^j}{\partial \tau_t} f_t^j}{\mu_t \sum_{j=1}^2 \frac{\partial \widetilde{d}_t^j}{\partial \tau_t} f_t^j}. \qquad (A.18)$$

Finally, using the definition of ζ_t^j, we get equation (**??**).

A.3 Proof of Proposition 3

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

$$\begin{aligned}
& \sum_{j=1}^2 \frac{\partial V_t^j}{\partial e_t} f_t^j G'\left(V_t^j\right)+\mu_t \sum_{j=1}^2 \tau_t \frac{\partial d_t^j}{\partial e_t} f_t^j+\lambda_t\left(\frac{\partial V_t^2}{\partial e_t}-\frac{\partial \widehat{V}_t^2}{\partial e_t}\right) \\
= \quad & -\upsilon_t \sum_{j=1}^2\left(\frac{\partial \pi^j}{\partial n_t^j} \frac{\partial n_t^j}{\partial e_t}+\frac{\partial \pi^j}{\partial e_t}\right) f_t^j-p'\left(e_t\right) \Upsilon_t, \qquad (A.19)
\end{aligned}$$

where, using the notation $q_t \equiv p(e_t) + \tau_t$, Υ_t is defined as:

$$\Upsilon_t \equiv \sum_{j=1}^{2} \frac{\partial V_t^j}{\partial q_t} f_t^j G'\left(V_t^j\right) + \mu_t \sum_{j=1}^{2} \tau_t \frac{\partial d_t^j}{\partial q_t} f_t^j + \lambda_t \left(\frac{\partial V_t^2}{\partial q_t} - \frac{\partial \widehat{V}_t^2}{\partial q_t} \right) + \upsilon_t \sum_{j=1}^{2} \frac{\partial \pi^j}{\partial n_t^j} \frac{\partial n_t^j}{\partial q_t} f_t^j. \quad (A.20)$$

The term $p'(e_t)\Upsilon_t$ 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 (A.19) as:

$$\left[f_t^1 G'\left(V_t^1\right) \frac{\partial V_t^1}{\partial B_t^1} - \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \right] \frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}} + \left[f_t^2 G'\left(V_t^2\right) + \lambda_t \right] \frac{\partial V_t^2}{\partial B_t^2} \frac{\frac{\partial V_t^2}{\partial e_t}}{\frac{\partial V_t^2}{\partial B_t^j}} + \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \left(\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}} - \frac{\frac{\partial \widehat{V}_t^2}{\partial e_t}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} \right) +$$
$$+ \mu_t \sum_{j=1}^{2} \tau_t \frac{\partial d_t^j}{\partial e_t} f_t^j = -\upsilon_t \sum_{j=1}^{2} \left(\frac{\partial \pi^j}{\partial n_t^j} \frac{\partial n_t^j}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t} \right) f_t^j - p'(e_t)\Upsilon_t. \quad (A.21)$$

We now use (A.4) and (A.8) to get expressions for, respectively, $f_t^1 G'\left(V_t^1\right) \frac{\partial V_t^1}{\partial B_t^1} - \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1}$ and $\left[f_t^2 G'\left(V_t^2\right) + \lambda_t \right] \frac{\partial V_t^2}{\partial B_t^2}$ and substitute them in (A.21). This gives:

$$\left[-\upsilon_t \frac{\partial \pi^1}{\partial n_t^1} \frac{\partial n_t^1}{\partial B_t^1} - \mu_t \left(-1 + \tau_t \frac{\partial d_t^1}{\partial B_t^1} \right) \right] f_t^1 \frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}} + \left[-\upsilon_t \frac{\partial \pi^2}{\partial n_t^2} \frac{\partial n_t^2}{\partial B_t^2} - \mu_t \left(-1 + \tau_t \frac{\partial d_t^2}{\partial B_t^2} \right) \right] f_t^2 \frac{\frac{\partial V_t^2}{\partial e_t}}{\frac{\partial V_t^2}{\partial B_t^j}} +$$
$$\mu_t \sum_{j=1}^{2} \tau_t \frac{\partial d_t^j}{\partial e_t} f_t^j + \lambda_t \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \left(\frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}} - \frac{\frac{\partial \widehat{V}_t^2}{\partial e_t}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} \right) = -\upsilon_t \sum_{j=1}^{2} f_t^j \left(\frac{\partial \pi^j}{\partial n_t^j} \frac{\partial n_t^j}{\partial e_t} + \frac{\partial \pi^j}{\partial e_t} \right) - p'(e_t)\Upsilon_t. \quad (A.22)$$

From (A.11) we can see that at an optimum $\Upsilon_t = -\mu_t \sum_{j=1}^{2} d_t^j f_t^j$ holds. Therefore, dividing by μ_t all terms in the previous equation and rearranging, we get:

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

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

$$\sum_{j=1}^{2} \frac{\frac{\partial V_t^j}{\partial e_t}}{\frac{\partial V_t^j}{\partial B_t^j}} f_t^j = -\frac{\upsilon_t}{\mu_t} \sum_{j=1}^{2} \left(\frac{d\pi^j}{de_t} \right)_{dV^j=0} f_t^j +$$

$$\frac{\lambda_t}{\mu_t} \frac{\partial \widehat{V}_t^2}{\partial B_t^1} \left(\frac{\frac{\partial \widehat{V}_t^2}{\partial e_t}}{\frac{\partial \widehat{V}_t^2}{\partial B_t^1}} - \frac{\frac{\partial V_t^1}{\partial e_t}}{\frac{\partial V_t^1}{\partial B_t^1}} \right) - \tau_t \sum_{j=1}^{2} \left(\frac{\partial d_t^j}{\partial e_t} \right)_{dV^j=0} f_t^j + p'\left(e_t\right) \sum_{j=1}^{2} d_t^j f_t^j. \tag{A.23}$$

Finally, using (**??**), we can rewrite (A.23) as (**??**).

B Sensitivity analysis

Changing the degree of aversion to inequality ψ, the differences in the optimal policies between our model and the standard one are qualitatively confirmed, as one can see in Table B.1.[23] Irrespective of the model we focus on, an increase in the aversion to inequality from $\psi = 0.99$ to $\psi = -4$ calls for a larger subsidy to day care and for a higher (lower) marginal effective tax rate on low(high)-skilled agents.[24] The welfare loss does not change significantly with the aversion to inequality.

As to the parameter ν which captures the degree of complementarity/substitutability between parental time and day care in the quality of the early childhood environment, it clearly does not play any role in the standard model where the probability π^j is a constant. In our model, the value of $METR^1$ is almost unaffected by a change in the degree of complementarity, whereas $METR^2$ is higher and the subsidy on day care τ^d generally lower when complementarity is larger, as one can see in Table B.1. These results can be intuitively understood considering that, at the optimum, the time that parents spend with their children is lower than the time allocated to day care: increasing the degree of complementarity between the two inputs makes parental time by high-skilled agents more valuable in terms of its influence on the probability of their children to become high skilled and both $METR^2$ and τ^d need to be optimally changed in order to discourage the use of day care.

As to the welfare loss, it does not change significantly when ν goes from the benchmark value of 0.98 (elasticity of substitution equal to 50) to 0.8 (elasticity of substitution equal to 5).

[23] The only exception is represented by the values of $METR^1$ when the degree of risk aversion ψ is set equal to 0.99 (very close to the Utilitarian case in which $\psi = 1$). In this case, the effective marginal tax rate on low-skilled agents is negative, the more so in the case where the quality is optimally chosen.

[24] Clearly, in the standard model the marginal effective tax rate on the high skilled is always 0.

Table B.1: Optimal policies: sensitivity analysis

		Standard	Quality fixed	Quality optimized
$\psi = -4$	e	1.09	1.09	1.5
	T'^1	8.92%	8.53%	14.37%
	T'^2	0.59%	2.96%	1.68%
	$METR^1$	7.87%	7.57%	12.33%
	$METR^2$	0.00%	2.42%	0.56%
	τ_d	-42.33%	-37.62%	-58.54%
$\psi = -0.1$	e	1.09	1.09	1.5
	T'^1	4.01%	3.73%	6.53%
	T'^2	0.33%	3.47%	1.57%
	$METR^1$	3.40%	3.29%	4.91%
	$METR^2$	0.00%	3.23%	0.71%
	τ_d	-21.43%	-14.60%	-39.35%
$\psi = 0.99$	e	1.09	1.09	1.5
	T'^1	0.05%	-0.36%	-0.16%
	T'^2	0.01%	3.78%	1.26%
	$METR^1$	0.04%	-0.03%	-0.94%
	$METR^2$	0.00%	3.95%	0.87%
	τ_d	-0.30%	9.42%	-15.32%
$\nu = 0.9$	e	1.09	1.09	1.5
	T'^1	7.38%	6.82%	11.29%
	T'^2	0.52%	5.04%	4.25%
	$METR^1$	6.44%	6.29%	9.64%
	$METR^2$	0.00%	4.75%	3.41%
	τ_d	-36.28%	-18.06%	-38.08%
$\nu = 0.8$	e	1.09	1.09	1.5
	T'^1	7.38%	6.54%	10.58%
	T'^2	0.52%	7.87%	8.23%
	$METR^1$	6.44%	6.47%	9.46%
	$METR^2$	0.00%	7.84%	7.69%
	τ_d	-36.28%	-1.94%	-20.80%

Table B.2: Welfare loss: sensitivity analysis

	Quality fixed	Quality optimized
$\psi = -4$	3.83%	4.11%
$\psi = -0.1$	3.62%	4.16%
$\psi = 0.99$	3.48%	4.00%
$\nu = 0.9$	3.58%	3.79%
$\nu = 0.8$	3.43%	3.57%