CESifo WORKING PAPERS

6533
2017

June 2017

Child Care Subsidies, Quality, and Optimal Income Taxation

Spencer Bastani, Sören Blomquist, Luca Micheletto

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6533*
Category 1: Public Finance

Child Care Subsidies, Quality, and Optimal Income Taxation

Abstract

In this paper we examine the desirability of subsidizing child care expenditures in a model where parents can choose both the quantity and the quality of child care services they purchase in the market. Our vehicle of analysis is a Mirrleesian optimal tax framework where child care services not only enable parents to work, but also contribute to children's formation of human capital. In addition, there are externalities related to the parents' choice of child care arrangements for their offspring. Using a quantitative simulation model calibrated to the US economy, we evaluate the relative merits of some the most common forms of child care subsidies (tax deductions, tax credits, and opting-out public provision schemes) in terms of their effectiveness in alleviating the distortions associated with income taxation and increasing the quality of child care chosen by parents.

JEL-Codes: H210, H410.

Keywords: optimal income taxation, child care subsidies, tax deductibility, tax credit, public provision of private goods.

Spencer Bastani
Department of Economics and Statistics
Linnaeus University
Växjö / Sweden
spencer.bastani@lnu.de

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

Luca Micheletto
Department of Law
University of Milan / Italy
luca.micheletto@unimi.it

First version: April, 2012, this version: June 15, 2017
This is a substantially revised and extended version of Bastani et al. (2012), which circulated under the title "Child Care Subsidies, Quality and the Optimal Income Tax". We are grateful to Dan Anderberg, Robin Boadway, Bradley Heim, Wojciech Kopczuk, Pierre Pestieau, Panu Poutvaara, and Andrew Shephard as well as numerous seminar, workshop and conference participants for comments on earlier drafts. Financial support from the Jan Wallander and Tom Hedelius Foundation and the Nordic Tax Research Council is gratefully acknowledged.

1 Introduction

Public involvement in child care provision is a widespread phenomenon but is also a hotly debated issue. Rosen (1996) strongly advocates against public subsidies. Becker and Posner (2005) express a similar view and argue that the provision of child care services is best left to the market without public intervention. An opposite view is held by Currie (2006) and Waldfogel (2006), who provide a number of arguments for publicly subsidized child care. The OECD has taken a strong interest in the subject with several projects on child care and five large reports (OECD 2001, 2006, 2011, 2012, 2015), and it has also given some member countries strong advice to increase public funding of child care. In the US, one third of the costs of child care for children under 6 are paid for by government subsidies (Blau and Currie, 2006). In the Nordic countries publicly financed child care is remarkably widespread; in Sweden, for instance, 85% of children in ages 2-5 are in child care,[1] and to a substantial part (80-85%) this care is publicly financed.

Several arguments have been proposed to justify subsidizing child care expenditures. One claim that is often made is that child care subsidies are desirable since they make it possible for both parents to work.[2] However, even if it is the case that child care services are needed in order to allow for both parents to work, it does not necessarily follow that child care expenditures should be subsidized unless it is a goal in itself that both parents should work, for instance to promote gender equality. Moreover, alternative policy instruments, such as for instance EITC schemes or gender-based taxes, might be more effective to achieve this goal.

Another argument in favor of child care subsidies is that they represent a means to increase fertility and improve the demographic composition.[3]

Perhaps most importantly, child care outside the home may serve the purpose of improving child outcomes, in particular for children with a poor social background.[4] It is implicit in this argument that many parents either do not have the financial means to buy good quality child care or lack the knowledge of the benefits descending from high quality child care, and that therefore public intervention is needed. The OECD 2006 report pushes a similar argument and argues that child care has public good aspects; the human capital formation in children is not only of interest for the parents but for society as a whole. Moreover, in light of recent evidence that the rate of return to society from investment in the human capital of children is larger the

[1] Calculated from tables 62 and 500 in Statistisk Årsbok 2003.

[2] This is also a major argument given in the 2006 OECD report. It is studied by, among others, Blau and Robins (1988), Gustafsson and Stafford (1992), Ribar (1995) and Powell (2002).

[3] See, for example, OECD (2011).

[4] See, for instance, Blau and Currie (2006), Currie (2006) and Waldfogel (2006). More recently, Havnes and Mogstad (2011) have provided evidence on the long-run effects of child care. Using Norwegian data they find that subsidized child care has large positive effects on children's adult outcomes measured in their early 30s in terms of education, labor market attachment and welfare dependency. In a companion paper, Havnes and Mogstad (2015) report that there is a large heterogeneity in effects; the positive effects are particularly large for children from families below median levels of income. The benefits of subsidized child care are however not undisputed. For instance, Baker et al. (2008) find a negative short run effect of child care on children's noncognitive development.

younger the children are (as argued by, for instance, Carneiro and Heckman, 2003), it seems that public funding of child care could be even more important than public funding of primary education.

If one takes the benefits of child care as given, and decides that child care should be subsidized, it would seem natural to argue that the gains from subsidizing child care expenditures ought to be traded-off with the deadweight losses of the taxes that are needed to finance the subsidies. In a sense, however, the optimal tax literature has pushed the opposite argument, namely, that subsidies to child care have the potential to increase the efficiency of the tax system. The argument is related to the well-known result in the optimal tax literature that goods complementary to labor supply should be subsidized, and dates back to Corlett and Hague (1953). The logic is most easily understood in the context of the modern Mirrleesian optimal income tax model where market productivity is private information (and is the only source of heterogeneity among agents) and the government optimizes a nonlinear income tax. In this setting, to prevent high-skilled agents from "mimicking" low-skilled agents, i.e. lowering their labor supply to earn the same pre-tax income as a low-skilled person (and therefore benefit from a more favorable tax treatment), a downward distortion on the labor supply of low-skilled agents must be imposed. Suppose, in addition, that one hour of formal (purchased) child care is needed for every hour of market work and that the quality of child care, and therefore its hourly price, is fixed. Then, if a high-skilled person were to behave as a mimicker, he/she would purchase fewer hours of child care than a true low-skilled person (given that, earning a higher wage rate, he/she can produce the same pre-tax income as the low-skilled person with fewer hours of work), and therefore would spend less on child care. This implies that a child care subsidy would be more highly valued by a low-skilled person than by a high-skilled mimicker. Thus, by introducing a subsidy to child care expenditures while at the same time lowering the after-tax income of a low-skilled person (through an adjustment in the income tax), one can make mimicking less attractive without hurting a low-skilled person. This in turn would open for the possibility to alleviate the downward distortion on the labor supply of low-skilled agents, thereby increasing their labor supply and achieving redistribution at a lower efficiency cost.[5]

In this paper we analyze the desirability of child care subsidies in a Mirrleesian optimal income tax framework, focusing attention on two rationales for public support of child care expenditures: (i) that child care subsidies might enable the government to redistribute at a lower efficiency cost, and, (ii) that there are positive externalities associated with parents' choices of child care arrangements for their offspring.

A key observation is that the extent to which child care subsidies enable redistribution at a lower efficiency cost depends on the strength of the correlation between child care expenditures and labor supply. In the example that we discussed above, where one hour of child care is needed for every hour of work, this correlation is very strong. In fact, with this assumption, subsidies to child care are always welfare-improving since a high-skilled mimicker would

[5]This rationale for the subsidization of child care subsidies is emphasized in Blomquist et al. (2010).

always spend less on child care expenditure than the low-skilled person being mimicked. However, with a richer model determining child care expenditure this is not necessarily the case. An important contribution of our paper is to assess the role for child care expenditures through the lens of a calibrated model where both the quantity and quality of formal child care is a choice variable for parents. This opens up in particular for the possibility that a mimicker, even though he/she works fewer hours, may have higher child care expenditure if he/she chooses a child care arrangement of a higher quality. In addition to incorporating the quality dimension of child care, we employ a rich model of household time allocation. These aspects taken together imply that we evaluate the desirability of child care subsidies in a model with an empirically relevant correlation between child care expenditure and labor supply.

To capture the externality argument in favor of public support of child care expenditures, we assume that the quality of care that children receive, both at home and outside the home (in terms of formal care), affects the children's human capital formation. Moreover, we implicitly assume that the human capital formation is an important determinant of the productivity and future earnings of individuals. This implies that increased human capital investments have the potential to increase the future tax base, and for a given size of public expenditures, result in lower taxes for future generations. In addition, high quality child care can help the promotion of social skills, reduce rates of crime, teenage pregnancy, high school dropout rates, adverse health conditions and other social problems (Heckman 2006, Heckman and Masterov 2007). We allow for such externalities by taking a reduced form approach, incorporating into our social objective a term which depends positively on the human capital of all children in the economy. In this setting, public intervention may then be desirable as a means for internalizing externalities associated with parents' choices of child care arrangements. The externality term that we incorporate into the social welfare function is chosen in such a way that it helps identify those child care subsidy schemes that work particularly well as means to increase the human capital levels of children, especially those in low-income families.

Our model economy consists of two-parent families with the same (positive) number of children in child care ages, and where we allow households to differ in their market ability as well as in their nurturing ability. We assume that there is a fixed supply schedule for child care services, represented by a price function which determines the hourly price of center-based care as a function of its quality. Parents decide on their time allocation, dividing their time between market work, domestic child care and leisure. In addition, they decide on the quantity and quality of formal (center-based) child care for their children. The government uses an optimal nonlinear income tax, based on total household income, for redistributive purposes. In addition, the government can supplement the income tax with child care subsidies that may take the form of either tax deductions, refundable tax credits, or a public provision scheme. Our point of departure is that there is a special tax treatment for families with children in child care ages, an assumption that agrees with reality both in the US and in many other countries. Thus, in our analysis we keep the net tax revenue from this group of taxpayers constant.

The public provision scheme that we consider is of an opting-out kind and we study two variants of it. In one case agents can choose between getting center-based child care services of a given fixed quality (chosen by the government) at a subsidized price (in which case households are said to "opt-in"), and freely choosing their preferred quality of center-based services but bearing the full cost (in which case households are said to "opt-out"). Under the alternative public provision scheme the government provides for free child care services of a given fixed quality up to a given maximum amount of hours per household. If parents are not satisfied with the quality provided by the government, they can decide to opt out and freely choose their preferred quality for child care services, in which case they will have to pay the full cost. On the other hand, if parents decide to opt-in, they can have their kid at a child care center for how many hours they want, but they will have to pay for the hours in excess of those provided for free by the government. In this sense this second provision scheme combines features of an opting-out scheme and features of a topping-up scheme.

Our ultimate goal is to assess whether child care subsidies are indeed welfare-enhancing and, in the case they are, whether it is better to use tax deductions, tax credits, or a public provision scheme. An important point to emphasize is that a possible outcome of the analysis is that there should be no subsidies to child care, and that both the redistributive goals and the internalization of the externalities are best addressed by solely relying on the optimal nonlinear income tax. In fact, a fully nonlinear income tax is a flexible instrument which also includes the possibility of having an Earned Income Tax Credit (EITC). Moreover, even though previous studies in the optimal tax literature have highlighted that child care subsidies might reduce the efficiency costs of redistributing via an income tax, the result was obtained in models where quality of formal care was not a choice variable for parents and therefore child care expenditures were only a function of the number of hours spent by a child in a facility.[6]

To illustrate the main mechanisms underlying our analysis we first present a simplified two-type theoretical model where we describe the policy instruments that we consider and where we characterize the optimal design of the various child care subsidies that we allow for. We then proceed to construct a quantitative simulation model with an extended number of types calibrated to empirical wage distributions and time use patterns based on US data. In the quantitative analysis we analyze the optimal structure of taxes and subsidies and assess the welfare gains of alternative ways of subsidizing child care when the government is simultaneously optimizing a nonlinear income tax. As already mentioned, we focus on tax deductions and refundable tax credits, but we also consider the possibility of implementing two alternative public provision schemes.[7] For the case of tax credits and tax deductions we consider both

[6]See, for instance, Blomquist et al. (2010), Domeij and Klein (2013), Bastani et al. (2015), Ho and Pavoni (2016), Koehne and Sachs (2017), Guner et al. (2017).

[7]The study of the case where child care subsidies are administered through tax deductions or through an opting-out public provision scheme is a novelty of our analysis. With a few exceptions (such as Blomquist and Christiansen 1995), previous contributions in the optimal tax literature have either considered child care subsidies that are equivalent to refundable tax credits or subsidies administered through topping-up public provision schemes. In addition, their quantitative importance has not been assessed.

the "uniform" case when the fraction of child care expenditures that is deductible against the household earned income, or the fraction that can be credited against the gross income tax, is the same for all households, and the "means-tested" case when the fractions are allowed to be income-dependent.

Our results indicate that subsidizing child care expenditures by means of tax credits welfare-dominates subsidizing child care expenditures by means of tax deductions. However, we do not find these instruments to be efficient to relax binding incentive-compatibility (self-selection) constraints, which means that they cannot be justified as a way to achieve redistribution at lower efficiency costs. Intuitively, the reason is that tax credits and tax deductions cannot mitigate binding incentive-constraints due to the fact that when quality is a choice variable, it is much less likely that so-called mimickers, i.e. high skill individuals who reduce their labor supply in order to qualify for a more lenient tax treatment, have lower child care expenditure than households being mimicked, i.e. actual low skill households. Thus, with tax deductions or tax credits the desirability of child care subsidies comes entirely from the externalities related to parents' choice of child care arrangements for their offspring.

Turning attention to the public provision schemes, we find that both tax credit and tax deduction schemes are dominated by the opting-out public provision schemes. In addition, the public provision schemes deliver welfare gains even when externalities are absent and can therefore be justified also as a means to mitigate the distortions associated with income taxation. Intuitively, public provision schemes are more effective as instruments to enable redistribution at a lower efficiency cost as they offer a subsidy in combination with a fixed quality level, thereby making it much more likely that the mimicker has a lower expenditure than the household being mimicked.

The paper is organized as follows. In Section 2 we discuss in more details how our contribution relates to some recent papers analyzing the welfare consequences of child care subsidies. In Section 3 we set up a simplified two-type optimal tax model to derive an analytical characterization of the welfare properties of subsidizing child care expenditures by means of granting tax credits and/or tax deductions, or by using two alternative opting-out public provision schemes. In Section 4 we describe the empirically driven simulation approach that we employ to evaluate the social welfare effects of the subsidy schemes that we consider. Section 5 provides the quantitative results of our numerical simulations based on US data. Finally, Section 6 offers some concluding remarks.

2 Relation to the literature

In this paper we follow the Mirrleesian approach to optimal income taxation and ask the question whether child care subsidies, in the form of tax credits, tax deductions or public provision schemes can usefully supplement the nonlinear income tax in achieving redistribution and increasing the quality of the child care arrangements chosen by parents for their offspring. While the role of child care subsidies in promoting the quality of child care is a novelty in our setting, previous contributions have emphasized the relationship between child care subsidies and the efficiency of the tax system.

First and foremost, our paper relates to the Atkinson and Stiglitz (1976) theorem on the usefulness of commodity taxes in the presence of a general (nonlinear) labor income tax. According to that theorem, if the income tax is allowed to be nonlinear, commodity taxes are a redundant policy instrument when preferences are separable between leisure and other goods. If the separability condition is not satisfied, one should use commodity taxes and subsidies to discourage the consumption of goods/services that are substitutes with labor supply and encourage the consumption of goods/services that are complements with labor supply. Given that child care services are regarded as a primary example of goods that are complements with labor supply, a consequence of the Atkinson and Stiglitz (1976) theorem is that they should be subsidized or in any case be subject to a more lenient tax treatment (compared with other goods).[8] In relation to the Atkinson and Stiglitz result, we assess the efficiency-enhancing role for child care through the lends of a a model where the correlation between child care expenditure and hours of work arises endogenously through a rich model of household decision making where both the quantity and quality of child care are choice variables.

In an important contribution, Domeij and Klein (2013) study how child care subsidies can help achieving efficient labor wedges (both across time and across agents) in a dynamic Ramsey optimal tax problem. They recommend that child care expenditures should be made tax deductible. However, they do not consider a Mirleesian income tax setting, and differently from our paper, they disregard the quality dimension of child care services and assume that they are only needed when both parents work (one hour of child care is needed for every hour that both parents work).

In a recent paper, Guner et al. (2017) extend the analysis by Domeij and Klein (2013) in several directions and study the macroeconomic and welfare implications of transfers to households with children, including subsidies to child care. However, they do not consider a Mirrleesian optimal income tax setting, and even though the quantity of child care is a choice variable (and not strictly related to hours of work) and agents face different (exogenous) child care costs, child care quality is not a choice variable of agents.[9]

The paper most closely related to ours is Ho and Pavoni (2017), who also analyze the design of child care subsidies in a static Mirrleesian framework. In particular, they provide a rich set of results on the optimal manner in which to subsidize child care. However, even though their setting is similar to ours, they consider a different model of household decision-making; most importantly, they do not take into account the quality dimension of child care and do not analyze the same subsidy instruments as we do. Hence, we view their paper as complementary to ours.

[8]This is also the view expressed by Crawford et al. (2010) in one of the chapters contained in the Mirrlees Review (2010). See Bastani et al. (2015) and Koehne and Sachs (2017) for recent discussions of this result.

[9]See also Bick (2017) who employs a rich model of household behavior with fertility, labor force participation, and various child-care choices to study the welfare effects of two child care reforms in Germany.

Another paper that we view as complementary to our paper is Koehne and Sachs (2017) who do not consider welfare optima, but instead study Pareto-improving reforms in the form of tax-breaks for work-related goods (they use household services to exemplify their point).

In sum, even though the literature has provided several useful contributions analyzing the subsidization of child care in different settings, none of the previous studies allowed child care quality to be a choice variable, nor explored the externality-argument for subsidies to child care using a quantitative model. In addition, the above papers do not analyze the role of public provision schemes (which, as our analysis suggests, play a particular important role in models where the quality of formal child care is a choice variable for parents).

3 A simple theoretical model

In this section we present a simplified two-type version of the model that will later be used to quantitatively explore the welfare effects of alternative ways to subsidize child care. We consider a subpopulation of the economy, namely families with the same (positive) number of children in child care ages, implicitly assuming that the tax system can be tagged based on the number of children in child care ages living in a given household (agreeing with tax systems in many countries, including the US). For illustrative purposes, and given that the same kind of analysis can be carried out for each specific tagged group, we will assume that each household has one child in child care ages. This subgroup of agents consists of two types of households who differ in terms of market ability and nurturing ability. Assuming assortative mating and positive correlation between the two dimensions of ability, we let households of type 2 be the high-ability households, and households of type 1 the low-ability ones. Allowing for the possibility of a gender wage gap, we distinguish between w_f^i, the wage rate of the father in a type i ($i = 1, 2$) household, and w_m^i, the wage rate of the mother in a type i ($i = 1, 2$) household, and we assume $w_f^i \geq w_m^i$. Moreover, we also assume that $w_m^2 > w_f^1$.

The proportion of households of type i is denoted by π^i and the total number of households is normalized to unity. We also assume that in each household the child needs to be taken care of all the time. Care can either be provided by one of the parents (when they are not working in the market or they are not engaged in other activities without the child) or by means of external child care services offered by centers which differ in quality. Normalizing to one the time endowment of each of the two parents in a household, the time constraint for the mother can be written as:

$$L_m + h_m + \ell_m = 1, \tag{1}$$

where L_m represents the labor supply of the mother, h_m denotes the number of hours spent by the mother with the child, and ℓ_m represents leisure time spent by the mother without the child.

Similarly, the time constraint for the father can be written as:

$$L_f + h_f + \ell_f = 1. \tag{2}$$

Finally, the time constraint for the child is:

$$h_c + h_m + h_f = 1, \tag{3}$$

where h_c denotes the number of hours that the child spends in a child care facility.[10]

Parents derive disutility from labor, utility from time spent with the child and leisure time without the child, and from the consumption of a composite good denoted by c and treated as the numéraire of our economy. Moreover, parents are assumed to care about the overall quality of the child care arrangement for their child, the idea being that a higher quality of the early childhood environment fosters the human capital development of the child and ameliorates his/her future prospects as an adult. The overall quality of the child care arrangement depends on the time that parents spend with the child (h_f and h_m), the nurturing abilities of the parents (denoted by ω_m^i and ω_f^i for respectively the mother and the father in a type i household), the time that the child spends in a child care facility (h_c) and the quality of the chosen child care facility (denoted by q_c). The overall quality of the child care arrangement, denoted by q, can then be described as:

$$q = f\left(h_m, h_f, h_c, q_c; \omega_m^i, \omega_f^i\right), \tag{4}$$

where the distinction between the first four arguments of $f(\cdot)$ and the last two is that only the former represent choice variables for the parents.

In our analysis we will also leave open the possibility that there are positive externalities related to parents' choice of the overall quality of the child care arrangement for their offspring. This will be captured by an additively separable term in the social welfare function, which therefore does not affect the behavior of households. For this reason, we will only introduce the externality when we will present the government's tax design problem.

Assuming a unitary model of household decision making, agents' preferences are represented by the following household utility function:

$$U = v\left(c, h_m, h_f, \ell_m, \ell_f, L_m, L_f, q; w_m^i, w_f^i\right), \tag{5}$$

or, equivalently, taking into account the definition of q provided by (4) and the fact that, given the time constraints (1)-(3), only four variables in the set $\left\{h_m, h_f, \ell_m, \ell_f, L_m, L_f, h_c\right\}$ are truly

[10]Notice that in the time constraints (1)-(3) the possibility that parents spend joint time with their child is not contemplated. The reason is that even if we were to allow for a variable h_{fm} denoting hours spent together by both parents with their child, it would be optimal for the household to set $h_{fm} = 0$. Intuitively, setting $h_{fm} > 0$ would be suboptimal since it would be equivalent to a reduction in the total time endowment for the household as a whole.

unconstrained variables for the household:

$$U = v\left(c, h_m, L_m, L_f, h_c, q_c; w^i_m, w^i_f, \omega^i_m, \omega^i_f\right). \tag{6}$$

The laissez-faire hourly price of center-based child care services, denoted by p, is assumed to depend on the quality q_c of the child care facility through the increasing function $p = p(q_c)$. Total child care expenditures for a household are given by $D = p(q_c) h_c$.

Using $u^i\left(c, h_m, L_m, L_f, h_c, q_c\right)$ as a shorthand for (6), the problem solved by a type i household in the absence of government intervention can be described as follows:

$$\max_{h_m, L_m, L_f, h_c, q_c} \quad u^i\left(c, h_m, L_m, L_f, h_c, q_c\right)$$

subject to:

$$c = w^i_f L_f + w^i_m L_m - p(q_c) h_c, \tag{7}$$

where constraint (7) represents the household budget constraint.

3.1 The government's problem

A benevolent government chooses its available tax instruments to maximize an inequality-averse social welfare function defined over the utilities of the various households. In line with the informational structure characterizing the Mirrleesian optimal tax literature, we assume that the government knows the distribution of types in the population but cannot observe "who is who", meaning that it can observe neither L^i_j nor w^i_j (with $j = m, f$), while it can observe $Y^i = w^i_f L_f + w^i_m L_m$, the household earned income. Thus, the government is prevented from imposing personalized lump-sum taxes/transfers. Another variable which is assumed to be unobservable by the government is the quality of center-based child care chosen by agents (and therefore the unitary price of the child care services chosen by any given household). Instead, we assume that the government can observe the household expenditures on child care services.

The policy instruments at the government's disposal include a general (nonlinear) tax on (household) labor income and the possibility to subsidize child care expenditures in various ways. Given that child care expenditures D are assumed to be observable at the household level, the government could in principle design a general, non-separable, tax schedule where the household tax liability is an unrestricted joint function of Y and D: $T = T(Y, D)$. However, since real-world governments do not usually adopt such sophisticated tax schemes, and since our analysis has the ambition of being policy relevant, we will restrict attention to subsidy schemes which are commonly adopted in real economies. In particular, we will focus on:

a) tax deductibility;

b) refundable tax credits;

c) opting-out public provision schemes.

In the following subsection we discuss and formalize the government's problem when nonlinear income taxation is supplemented by tax deductibility and/or tax credit granted for child care expenditures. In a subsequent subsection we consider the government's problem when an optimal nonlinear income tax is supplemented with an opting-out public provision scheme.

3.1.1 The government's problem with tax deductibility and/or tax credits

Let α and β denote respectively the fraction of child care expenditures that can be deducted against the household earned income and the fraction that can be credited against the gross income tax. Taxable income, denoted by M, can then be defined as $M = w_f^i L_f + w_m^i L_m - \alpha D$. The gross income tax is obtained by applying the tax schedule $T(M)$ and the net income tax is obtained by subtracting βD from $T(M)$. Using the notation $B \equiv M - T(M)$, and assuming that the tax credit is refundable, the household consumption can be expressed as:

$$\begin{aligned} c &= w_f^i L_f + w_m^i L_m - T(M) - (1-\beta) D \\ &= w_f^i L_f + w_m^i L_m - M + B - (1-\beta) D \\ &= w_f^i L_f + w_m^i L_m - w_f^i L_f - w_m^i L_m + \alpha D + B - (1-\beta) D \\ &= B - (1-\alpha-\beta) D. \end{aligned}$$

With two household types, the problem of optimally choosing the nonlinear tax function $T(M)$ can be equivalently stated as the problem of selecting two bundles in the (M, B)-space, one for each household type, subject to a public budget constraint and subject to the self-selection constraints requiring that each household type (weakly) prefers the (M, B)-bundle intended for it by the government to that intended for the other type. Hereafter, we will use the term "mimicker" to refer to a household misrepresenting its true type to the government by choosing the bundle intended for another type.

Before formally presenting the government's problem when a nonlinear income tax is supplemented by tax deductibility and/or tax credit for child care expenditures, we need to introduce some more notation. For this purpose, we denote by $V^i(M, B, \alpha, \beta)$ the conditional indirect utility obtained by an optimizing household of type i at a given (M, B)-bundle for given values of α and β. Substituting $\left[M - w_m^i L_m + \alpha p(q_c) h_c\right]/w_f^i$ for L_f (for a given value of M, once q_c, h_c and L_m have been chosen, L_f is residually given), we have:

$$V^i(M, B, \alpha, \beta) = \max_{q_c, h_c, h_m, L_m} u^i\left(B - (1-\alpha-\beta) p(q_c) h_c, h_m, L_m, \frac{M - w_m^i L_m + \alpha p(q_c) h_c}{w_f^i}, h_c, q_c\right).$$

The solution to the problem above yields the conditional demands $q_c^i \equiv q_c(\cdot), h_c^i \equiv h_c(\cdot), h_m^i \equiv h_m(\cdot)$, and $L_m^i \equiv L_m(\cdot)$ as functions of $M, B, \alpha, \beta; w_m^i, w_f^i, \omega_m^i, \omega_f^i$.

The indirect utility for a type i household is then obtained by optimally choosing the (M, B)-

bundle among those implied by the nonlinear schedule $B = M - T(M)$. From this optimization problem, which can be formalized as $\max_M V^i(M, M - T(M), \alpha, \beta)$, one can define the implicit marginal income tax rate faced by a parent in household i as:

$$T' = 1 + \frac{\partial V^i/\partial M}{\partial V^i/\partial B} = 1 - MRS^i_{MB}, \tag{8}$$

where $-\frac{\partial V^i/\partial M}{\partial V^i/\partial B} = MRS^i_{MB}$ represents the marginal rate of substitution between M and B for a household of type i at the selected (M, B)-bundle.

Applying the envelope theorem one gets:

$$\frac{\partial V^i}{\partial M} = \frac{1}{w^i_f}\frac{\partial u^i}{\partial L_f}, \tag{9}$$

$$\frac{\partial V^i}{\partial B} = \frac{\partial u^i}{\partial c}, \tag{10}$$

$$\frac{\partial V^i}{\partial \alpha} = p(q_c)h_c\frac{\partial u^i}{\partial c} + \frac{p(q_c)h_c}{w^i_f}\frac{\partial u^i}{\partial L_f} = p(q_c)h_c\left[1 + \frac{1}{w^i_f}\frac{\partial u^i}{\partial L_f}/\frac{\partial u^i}{\partial c}\right]\frac{\partial u^i}{\partial c} = p(q_c)h_cT'\frac{\partial u^i}{\partial c}, \tag{11}$$

$$\frac{\partial V^i}{\partial \beta} = p(q_c)h_c\frac{\partial u^i}{\partial c}, \tag{12}$$

$$\frac{\partial V^i}{\partial p} = \left[-(1-\alpha-\beta)\frac{\partial u^i}{\partial c} + \frac{\alpha}{w^i_f}\frac{\partial u^i}{\partial L_f}\right]h_c. \tag{13}$$

Equation (11) shows that the private welfare gain of a marginal increase in the deductibility rate is proportional to the product between the marginal income tax rate faced by an agent and the expenditure on child care services.

Equation (12) shows that the private welfare gain of a marginal increase in the fraction of child care expenditures which can be claimed as a refundable tax credit is proportional to the private expenditure on child care services.

Equation (13) determines an upper bound for the values of α and β. The reason is that we need to make sure that $\partial V^i/\partial p \leq 0$; otherwise, the household consumption and utility would be increasing in the price of child care services. This would imply that the seller and the buyer would both be benefitting from a higher price, and the possibilities for "cheating" or colluding against the government would be large. To ensure that $\partial V^i/\partial p \leq 0$ it must be that $\frac{\alpha}{w^i_f}\frac{\partial u^i}{\partial L_f} \leq (1-\alpha-\beta)\frac{\partial u^i}{\partial c}$. When $\alpha = 0$ (no tax deductibility) the condition is satisfied provided that $\beta \leq 1$. When $\alpha > 0$, the condition is satisfied when $(1-\alpha-\beta)/\alpha \geq \frac{\partial u^i}{\partial L_f}/\left(w^i_f\frac{\partial u^i}{\partial c}\right)$, which is in turn equivalent to requiring that $(1-\beta)/\alpha \geq 1 + \frac{\partial u^i}{\partial L_f}/\left(w^i_f\frac{\partial u^i}{\partial c}\right)$. Given the definition that we have provided above for the implicit marginal income tax rate faced by a member of household i, the previous inequality can also be rewritten as $T' \leq (1-\beta)/\alpha$ or, equivalently, $\alpha T' + \beta \leq 1$. This condition has to be satisfied for all households, including the mimickers at the self-selection constraints which are binding at the solution to the government's problem. Given that T' cannot exceed unity, a simple sufficient condition which ensures that $\partial V^i/\partial p \leq 0$ is to assume right from the outset that $\alpha + \beta \leq 1$. Since this simpler condition avoids some complexities which are especially relevant when numerically simulating the model, we will hereafter impose this constraint on our analysis.

We are now ready to formally state the government's problem. Using θ^i to denote the welfare weight applied by the government on the utility of households of type i (with $\theta^1 + \theta^2 = 1$), and denoting by $v\left(q^1, q^2\right)$ the additively separable term that captures the externalities related to parents' choice of the overall quality of child care arrangements for their offspring, the government's problem can be described as:

$$\max_{M^1,B^1,M^2,B^2,\alpha,\beta} \left\{\theta^1 V^1\left(M^1, B^1, \alpha, \beta\right) + \theta^2 V^2\left(M^2, B^2, \alpha, \beta\right) + v\left(q^1, q^2\right)\right\},$$

subject to:

$$V^2\left(M^2, B^2, \alpha, \beta\right) \geq V^2\left(M^1, B^1, \alpha, \beta\right), \qquad (\lambda)$$

$$\sum_{i=1}^{2}\left(M^i - B^i\right)\pi^i \geq \beta \sum_{i=1}^{2} \pi^i p\left(q_c^i\right) h_c^i + R, \qquad (\mu)$$

$$1 \geq \alpha + \beta, \qquad (\delta)$$

where Lagrange multipliers are within parentheses. The λ -constraint represents the self-selection constraint requiring a high-skilled household not to mimic a low-skilled household,[11] and the μ-constraint represents the public budget constraint with R denoting the exogenous value of the net tax (if positive) or transfer (if negative) that the government wants to collect from (or transfer to) the set of households with one child in child care ages.[12]

The first order condition of this government's problem are presented in appendix A. Since our primary interest in this subsection is to characterize the optimal design of a tax credit and/or a tax deductibility policy, we also relegate to the appendix the formulas defining the optimal marginal income tax rates faced by the different groups of households. Using a " $\frown$ " symbol to denote a variable when pertaining to a mimicker and a tilde symbol to denote compensated (Hicksian) demands, the following Proposition characterizes the optimal values for the policy parameters α and β.

Proposition 1. *Denoting by $\widehat{T'}\left(M^1\right)$ the implicit marginal income tax rate faced by a mimicker at the bundle intended by the government for low-skilled agents, $\widehat{T'}\left(M^1\right) \equiv 1 + \frac{\partial \widehat{V^2}/\partial M^1}{\partial \widehat{V^2}/\partial B^1}$, the*

[11] Since we confine attention to the case when redistribution goes from high- to low-skilled households, we can safely disregard the self-selection constraint requiring low-skilled households not to be tempted to mimic high-skilled households.

[12] It should be kept in mind that we are solving for the optimal policy that applies to a specific group of households in the society, those with one child in child care ages, assuming that the government can tag the income tax system based on the number of children in child care ages that live in a household.

optimal deductibility rate for child care expenditures is characterized by the condition:

$$\sum_{i=1}^{2}\pi^{i}\left[\beta\frac{d\widetilde{D}^{i}}{d\alpha}-D^{i}T'\left(M^{i}\right)\right] = \frac{\lambda\frac{\partial\widehat{V}^{2}}{\partial B^{1}}\left[D^{1}T'\left(M^{1}\right)-\widehat{D}^{2}\widehat{T}'\left(M^{1}\right)\right]+\sum_{i=1}^{2}\frac{\partial v\left(q^{1},q^{2}\right)}{\partial q^{i}}\frac{d\widetilde{q}^{i}}{d\alpha}-\delta}{\mu}; \quad (14)$$

the optimal fraction of child care expenditures which can be claimed as a refundable tax credit is instead given by:

$$\beta=\frac{\lambda\frac{\partial\widehat{V}^{2}}{\partial B^{1}}\left(D^{1}-\widehat{D}^{2}\right)+\sum_{i=1}^{2}\frac{\partial v\left(q^{1},q^{2}\right)}{\partial q^{i}}\frac{d\widetilde{q}^{i}}{d\beta}-\delta}{\mu\sum_{i=1}^{2}\pi^{i}d\widetilde{D}^{i}/d\beta}. \quad (15)$$

Proof See Appendix A.1. □

To interpret the result stated in the first part of Proposition 1 it is useful to think of the following policy experiment. Starting from an initial equilibrium, raise marginally the deductibility rate while at the same time adjusting B^i by $dB^i = -MRS^{\,i}_{\alpha B}d\alpha = -\frac{\partial V^i/\partial\alpha}{\partial V^i/\partial B^i}d\alpha$ in order to offset any welfare effects on non-mimicking households. By construction, the only possible effects of the reform are those on the self-selection constraint, the government's budget constraint, the externality term entering the social objective function, and the constraint defining an upper bound for α. The first effect is captured by the first term at the numerator of the right hand side of (14). From (10) and (11) one realizes that $D^1T'\left(M^1\right)-\widehat{D}^2\widehat{T}'\left(M^1\right)$ represents the difference between how much the government can lower B^1 without reducing the utility of a low-skilled household and the reduction in B^1 which would be required to maintain the mimicker's utility unaffected. When $D^1T'\left(M^1\right)-\widehat{D}^2\widehat{T}'\left(M^1\right)>0$, the reform makes the mimicker worse-off and in this way it slackens the self-selection constraint.

The term on the left hand side of (14) captures the effects of the reform on the public budget. On one hand, thanks to the reduction in B^i, the reform allows the government to raise the tax revenue by $\sum_{i=1}^{2}\pi^iD^iT'\left(M^i\right)$. On the other hand, total outlays for the government might increase if private child care expenses are also eligible for a tax credit treatment ($\beta>0$) and the parents' behavioral response to the reform is to increase private expenses on child care services (with the behavioral response captured by $\sum_{i=1}^{2}\pi^i\beta\frac{d\widetilde{D}^i}{d\alpha}$).

The third term on the right hand side of (14) reflects the Pigouvian motives for allowing child care expenditures to be tax deductible: raising α offers an additional source of welfare gains if there are positive externalities associated with an increase in the overall quality of child care arrangement chosen by parents and an increase in α induces parents to change their behavior in such a way to increase q.

Finally, the last term on the right hand side of (14) captures the effect of the reform on the constraint defining an upper bound for α. This term vanishes when the δ-constraint is not binding. Otherwise, the compensated increase in the deductibility rate implies an additional cost for the government.

Denoting by $\mathcal{L}$ the Lagrangian of the government's problem, in the event that there is no tax credit granted for child care expenses one can show that, when the income tax $T(M)$ is optimally chosen, $\partial\mathcal{L}/\partial\alpha$ can be expressed as:[13]

$$\frac{\partial\mathcal{L}}{\partial\alpha}=\lambda\frac{\partial\widehat{V}^2}{\partial B^1}\left[D^1-\widehat{D}^2\right]\widehat{T}'\left(M^1\right)-\delta+\sum_{i=1}^{2}\frac{\partial v\left(q^1,q^2\right)}{\partial q^i}\left[\frac{d\widetilde{q}^i}{d\alpha}-D^i\left(\frac{dq^i}{dM^i}\right)_{dV^i=0}\right]. \tag{16}$$

In the absence of any externality related to the overall quality of the child care arrangement chosen by parents, (16) reduces to:

$$\frac{\partial\mathcal{L}}{\partial\alpha}=\lambda\frac{\partial\widehat{V}^2}{\partial B^1}\left[D^1-\widehat{D}^2\right]\widehat{T}'\left(M^1\right)-\delta,$$

showing that, when tax deductibility is used in isolation, the deductibility rate should either be pushed to the maximum level ($\alpha=1$), which happens if the child care expenses of a low-skilled household exceed those of a mimicker ($D^1>\widehat{D}^2$), or should be set to zero if $D^1<\widehat{D}^2$.

Eq. (15) admits an interpretation which is similar to the one that applies to the optimal commodity tax rules in models where a set of linear commodity taxes supplements a nonlinear income tax. On one hand it embeds a trade-off between the mimicking-deterring effects which are produced, when $D^1>\widehat{D}^2$, by a compensated increase in β (namely a marginal increase in β which is accompanied by adjusting B^i by $dB^i=-MRS^i_{\beta B}d\beta=-\frac{\partial V^i/\partial\beta}{\partial V^i/\partial B^i}d\beta$ to prevent any welfare effect on non-mimicking households), and the deadweight losses arising from distorting the parents' behavior ($\sum_{i=1}^{2}\pi^i\frac{d\widetilde{D}^i}{d\beta}>0$). On the other hand, it also accounts for the Pigouvian motives to affect the parents' choice about the overall quality of the child care arrangement for their offspring.

3.1.2 The government's problem with an opting-out public provision scheme

In this section we consider the possibility that child care services are publicly provided by means of an opting-out provision scheme.[14] This means that the government provides child care services of a given fixed quality $\overline{q}_c$ at a subsidized cost for parents. If parents are not satisfied with the quality provided by the government, they can decide to opt out and freely choose their preferred quality for child care services; however, if they do it, they have to pay the full cost.

If an opting-out provision scheme allows the government to increase social welfare beyond the value characterizing a pure income tax optimum, it could either be the case that only type 1 households opt-in or that both households opt-in. Since in a model with several different types

[13] The details of the derivation are provided in appendix A.

[14] The desirability of using in-kind transfers as redistributive devices was first discussed by Nichols and Zeckhauser (1982), Guesnerie and Roberts (1984), Blackorby and Donaldson (1988) and Besley and Coate (1992). In the context of the optimal tax literature, public provision schemes have been analyzed in Blomquist and Christiansen (1995, 1998), Boadway and Marchand (1995), Cremer and Gahvari (1997), Gahvari and Mattos (2007), Blomquist et al. (2010, 2016).

it is likely that some households will opt-out at a social optimum, here we restrict attention to the case when the government finds it optimal that only type 1 families opt-in.

Denote by s the subsidy rate for households who opt-in. The hourly co-payment for households who opt in is then $(1-s)\,p(\overline{q}_c)$. Denoting by V^{iin} the indirect utility for a type-i household who opts in and by V^{iout} the indirect utility for a type-i household who opts out, we have:

$$V^{iin}(M,B,\overline{q}_c,s) = \max_{h_c,h_m,L_m} u^i\left(B-(1-s)\,p(\overline{q}_c)\,h_c, h_m, L_m, \frac{M-w^i_m L_m}{w^i_f}, h_c, \overline{q}_c\right);$$
$$V^{iout}(M,B) = \max_{q_c,h_c,h_m,L_m} u^i\left(B-p(q_c)\,h_c, h_m, L_m, \frac{M-w^i_m L_m}{w^i_f}, h_c, q_c\right).$$

Denoting by h_c^{1in} the hours of center-based care demanded by low-skilled households who opt-in, the government's problem can then be formally stated as follows:

$$\max_{M^1,B^1,M^2,B^2,\overline{q}_c,s} \left\{\theta^1 V^{1in}\left(M^1,B^1,\overline{q}_c,s\right)+\theta^2 V^{2out}\left(M^2,B^2\right)+v\left(q^1,q^2\right)\right\},$$

subject to:

$$\begin{aligned}
V^{1in}\left(M^1,B^1,\overline{q}_c,s\right) &\geq V^{1out}\left(M^1,B^1\right) && \left(\gamma^1\right)\\
V^{2out}\left(M^2,B^2\right) &\geq V^{2in}\left(M^2,B^2,\overline{q}_c,s\right) && \left(\gamma^2\right)\\
V^{2out}\left(M^2,B^2\right) &\geq V^{2out}\left(M^1,B^1\right) && \left(\lambda^{out}\right)\\
V^{2out}\left(M^2,B^2\right) &\geq V^{2in}\left(M^1,B^1,\overline{q}_c,s\right) && \left(\lambda^{in}\right)\\
\sum_{i=1}^{2}\left(M^i-B^i\right)\pi^i &\geq \pi^1 sp(\overline{q}_c)\,h_c^{1in}+R && (\mu)
\end{aligned}$$

In the problem above the γ-constraints represent the individual rationality constraints. The constraint with associated multiplier γ^1 requires that a low-skilled household is better off by opting-in. Similarly, the constraint with associated multiplier γ^2 requires that, when choosing the (M^2,B^2)-bundle intended for it by the government, a high-skilled household is better off by opting-out. The λ-constraints represent the self-selection constraints associated with the two possible deviating strategies available to a high-skilled household: either to choose the (M^1,B^1)-bundle and opt-out (λ^{out}-constraint) or choose the (M^1,B^1)-bundle and opt-in (λ^{in}-constraint).

Using a " $\frown$ " symbol to denote a variable when pertaining to a mimicker and a tilde symbol to denote compensated (Hicksian) demands, the following Proposition characterizes the optimal values for the policy variables $\overline{q}_c$ and s.

Proposition 2. *Denoting by $MRS^{1in}_{\overline{q}_c B}$ and $\widehat{MRS}^{2in}_{\overline{q}_c B}$ the marginal rate of substitution between $\overline{q}_c$ and B for, respectively, an opting-in type-1 household and an opting-in type-2 mimicker, the optimal quality set by the government for the publicly provided child care services abides by* *the following condition:*

$$MRS^{1in}_{\overline{q}_c B} = s\left[h^{1in}_c p'\left(\overline{q}_c\right) + p\left(\overline{q}_c\right)\left(\frac{\partial h^{1in}_c}{\partial \overline{q}_c}\right)_{dV^{1in}=0}\right] - \frac{\partial v\left(q^1, q^2\right)/\partial q^1}{\mu\pi^1}\left(\frac{dq^1}{d\overline{q}_c}\right)_{dV^{1in}=0}$$
$$+\frac{\gamma^2}{\mu\pi^1}\frac{\partial V^{2in}}{\partial \overline{q}_c} - \frac{\gamma^1}{\mu\pi^1}\frac{\partial V^{1out}}{\partial B^1}MRS^{1in}_{\overline{q}_c B}$$
$$+\frac{\lambda^{in}}{\mu\pi^1}\frac{\partial \widehat{V}^{2in}}{\partial B^1}\left(\widehat{MRS}^{2in}_{\overline{q}_c B} - MRS^{1in}_{\overline{q}_c B}\right) - \frac{\lambda^{out}}{\mu\pi^1}\frac{\partial \widehat{V}^{2out}}{\partial B^1}MRS^{1in}_{\overline{q}_c B}. \tag{17}$$

At an optimum, the government subsidizes the child care expenditures for the opting-in households at a rate s that satisfies the following condition:

$$s = \left[\gamma^1\frac{\partial V^{1out}}{\partial B^1}h^{1in}_c - \gamma^2 h^{2in}_c\frac{\partial V^{2in}}{\partial B^2}\right]\frac{1}{\mu\pi^1\partial\widetilde{h}^{1in}_c/\partial s} \tag{18}$$
$$+\left[\lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial B^1}\left(h^{1in}_c - \widehat{h}^{2in}_c\right) + \lambda^{out}\frac{\partial \widehat{V}^{2out}}{\partial B^1}h^{1in}_c\right]\frac{1}{\mu\pi^1\partial\widetilde{h}^{1in}_c/\partial s} + \frac{\partial v\left(q^1, q^2\right)/\partial q^1}{\mu\pi^1 p\left(\overline{q}_c\right)\partial\widetilde{h}^{1in}_c/\partial s}\left(\frac{dq^1}{ds}\right)_{dV^{1in}=0}.$$

Proof See Appendix A.2. □

Condition (17) characterizes the optimal quality chosen by the government for the publicly provided child care services by means of a modified Samuelson-type condition where the left hand side represents the marginal willingness to pay of low-skilled households (the only agents who in equilibrium opt-in) for a marginal increase in $\overline{q}_c$.

On the right hand side, the first term captures the budget cost incurred by the government when marginally raising $\overline{q}_c$. This budget cost is given by the sum of a mechanical and a behavioral term, the former capturing the increased unitary cost of child care services being publicly provided and the latter the public budget effect coming from the change in the hours of center-based demanded by low-skilled households.

The second term on the right hand side captures the impact of a compensated (for low-skilled households) marginal increase in $\overline{q}_c$ on the externality-term entering the objective function maximized by the government. Provided that there are positive externalities associated with an increase in the overall quality of child care arrangement for kids in low-skilled households (i.e. an increase in q^1), the effective marginal cost of raising $\overline{q}_c$ is lowered if a (compensated) marginal increase in $\overline{q}_c$ induces an increases in q^1.

The third and fourth term on the right hand side of (17) capture the effects on the individual rationality constraints of a marginal increase in the quality of publicly provided care that is accompanied by a reduction in B^1 that leaves unaffected the utility of opting-in low-skilled households. On one hand such an increase in $\overline{q}_c$ makes more tempting for high-skilled households to opt-in (which represents a cost for the government, if it aims at having high-skilled households to opt-out). On the other hand, a compensated (for opting-in low-skilled households) marginal increase in the quality of the publicly provided care makes it also more attractive for low-skilled households to opt-in (which represent a benefit, from the point of view of the government, if it aims at having low-skilled households to opt-in).

The last two terms on the right hand side of (17) capture instead the mimicking-deterring effects associated with a compensated (for opting-in low-skilled households) marginal increase in the quality of the publicly provided care. On one hand, such a reform lowers the utility for a high-skilled household planning to choose the (M, B)-bundle intended for low-skilled households while at the same time opting-out (last term on the right hand side of (17)). On the other hand, whether the reform makes less attractive for high-skilled households to mimic and opt-in depends on the difference between $\widehat{MRS}^{2in}_{\overline{q}_c B}$ and $MRS^{1in}_{\overline{q}_c B}$. If this difference is negative (resp.: positive) this type of mimicking strategy becomes less (resp.: more) attractive, representing for the government an additional benefit (resp.: cost) of raising $\overline{q}_c$.

Condition (19) characterizes the optimal subsidy rate for publicly provided care by trading off the deadweight losses, in terms of increasingly distorting the choice of opting-in households regarding the hours of center-based care for their kids, with the potential gains in terms of mimicking-deterring effects, externality-internalizing effects, and effects on the individual rationality constraints. With respect to these last effects, a marginal increase in the subsidy rate s, accompanied by a reduction in B^1 that leaves unaffected the utility of opting-in low-skilled households, tightens the individual-rationality constraint requiring high-skilled households not to opt-in (γ^2-constraint), while making it easier to induce low-skilled households to opt-in (γ^1-constraint). Regarding the mimicking-deterring effects of a compensated (for opting-in low-skilled households) marginal increase in s, notice that mimicking and opting-out becomes less attractive for high-skilled households. On the other hand, whether it becomes less attractive for high-skilled households to mimic and opt-in depends on the sign of $h_c^{1in} - \widehat{h}_c^{2in}$. If this difference is positive (resp.: negative) this type of mimicking strategy becomes less (resp.: more) attractive, representing for the government an additional benefit (resp.: cost) of raising s.

3.1.3 The government's problem with an alternative opting-out public provision scheme

In this section we consider the possibility that child care services are publicly provided by means of a scheme that combines features of an opting-out scheme and features of a topping-up scheme. In particular, we will assume that the government provides for free child care services of a given fixed quality $\overline{q}_c$ up to a given maximum amount of hours, denoted by $\overline{h}_c$, per household. If parents are not satisfied with the quality provided by the government, they can decide to opt out and freely choose their preferred quality for child care services; however, if they do it, they have to pay the full cost. On the other hand, if parents decide to opt-in they can have their kid at a child care center for more than $\overline{h}_c$ hours, but then they will have to pay for the hours in excess of $\overline{h}_c$. In this sense the provision scheme combines features of an opting-out scheme and features of a topping-up scheme. It can be regarded as being of an opting-out type with respect to quality and of a topping-up type with respect to hours of child care. For simplicity, we will hereafter use the label "impure" opting-out scheme to refer to this alternative public provision scheme.

As for the case considered in the previous subsection, if an impure opting-out scheme allows the government to increase social welfare beyond the value characterizing a pure income tax optimum, it could either be the case that only type 1 households opt-in or that both households opt-in. Since in a model with several different types it is likely that some households will opt-out at a social optimum, here we restrict attention to the case when the government finds it optimal that only type 1 families opt-in.

Denoting by V^{iin} the indirect utility for a type-i household who opts in and by V^{iout} the indirect utility for a type-i household who opts out, we have:

$$V^{iin}\left(M, B, \overline{q}_c, \overline{h}_c\right) = \max_{h_c, h_m, L_m} u^i\left(B - p\left(\overline{q}_c\right)\left(h_c - \overline{h}_c\right), h_m, L_m, \frac{M - w_m^i L_m}{w_f^i}, h_c, \overline{q}_c\right);$$
$$V^{iout}\left(M, B\right) = \max_{q_c, h_c, h_m, L_m} u^i\left(B - p\left(q_c\right) h_c, h_m, L_m, \frac{M - w_m^i L_m}{w_f^i}, h_c, q_c\right).$$

Denoting by h_c^{1in} the hours of center-based care demanded by low-skilled households who opt-in, notice that, in a two-type model where only low-skilled agents opt-in at an optimum, the optimal policy chosen by the government will be such that $\overline{h}_c \leq h_c^{1in}$.[15] Thus, we can formally state the government's problem as follows:

$$\max_{M^1, B^1, M^2, B^2, \overline{q}_c, \overline{h}_c} \left\{\theta^1 V^{1in}\left(M^1, B^1, \overline{q}_c, \overline{h}_c\right) + \theta^2 V^{2out}\left(M^2, B^2\right) + \nu\left(q^1, q^2\right)\right\},$$

subject to:

$$\begin{aligned}
V^{1in}\left(M^1, B^1, \overline{q}_c, \overline{h}_c\right) &\geq V^{1out}\left(M^1, B^1\right) && \left(\gamma^1\right)\\
V^{2out}\left(M^2, B^2\right) &\geq V^{2in}\left(M^2, B^2, \overline{q}_c, \overline{h}_c\right) && \left(\gamma^2\right)\\
V^{2out}\left(M^2, B^2\right) &\geq V^{2out}\left(M^1, B^1\right) && \left(\lambda^{out}\right)\\
V^{2out}\left(M^2, B^2\right) &\geq V^{2in}\left(M^1, B^1, \overline{q}_c, \overline{h}_c\right) && \left(\lambda^{in}\right)\\
\sum_{i=1}^{2}\left(M^i - B^i\right)\pi^i &\geq \pi^1 p\left(\overline{q}_c\right)\overline{h}_c + R, && (\mu)
\end{aligned}$$

where the γ-constraints represent the individual rationality constraints, the λ-constraints represent the self-selection constraints associated with the two possible deviating strategies available to a high-skilled household, and the μ-constraint represents the public budget constraint.

Using a " $\frown$ " symbol to denote a variable when pertaining to a mimicker and a tilde symbol to denote compensated (Hicksian) demands, the following Proposition characterizes the optimal

[15]Given that households who opt-in can choose for how many hours to have their kid at a child care facility, there is no reason to set $\overline{h}_c > h_c^{1in}$.

values for the policy variables $\overline{q}_c$ and $\overline{h}_c$.

Proposition 3. *Denoting by $MRS^{1in}_{\overline{q}_c B}$ and $\widehat{MRS}^{2in}_{\overline{q}_c B}$ the marginal rate of substitution between $\overline{q}_c$ and B for, respectively, an opting-in type-1 household and an opting-in type-2 mimicker, the optimal quality set by the government for the publicly provided child care services abides by the following condition:*

$$\begin{aligned} MRS^{1in}_{\overline{q}_c B} &= \overline{h}_c p'(\overline{q}_c) - \frac{\partial v\left(q^1, q^2\right)/\partial q^1}{\mu\pi^1}\left(\frac{dq^1}{d\overline{q}_c}\right)_{dV^{1in}=0} \\ &+ \frac{\gamma^2}{\mu\pi^1}\frac{\partial V^{2in}}{\partial \overline{q}_c} - \frac{\gamma^1}{\mu\pi^1}\frac{\partial V^{1out}}{\partial B^1} MRS^{1in}_{\overline{q}_c B} \\ &+ \frac{\lambda^{in}}{\mu\pi^1}\frac{\partial \widehat{V}^{2in}}{\partial B^1}\left(\widehat{MRS}^{2in}_{\overline{q}_c B} - MRS^{1in}_{\overline{q}_c B}\right) - \frac{\lambda^{out}}{\mu\pi^1}\frac{\partial \widehat{V}^{2out}}{\partial B^1} MRS^{1in}_{\overline{q}_c B}. \end{aligned} \tag{19}$$

Denoting by $\widehat{MRS}^{2in}_{\overline{h}_c B}$ the marginal rate of substitution between $\overline{q}_c$ and B for an opting-in type-2 mimicker, the government provides for free child care services of quality $\overline{q}_c$ for a maximum amount of hours implicitly characterized by the following condition:

$$\begin{aligned} \frac{\partial v\left(q^1, q^2\right)/\partial q^1}{\mu\pi^1}\left(\frac{dq^1}{d\overline{h}_c}\right)_{dV^{1in}=0} &= \frac{\gamma^2}{\mu\pi^1}\frac{\partial V^{2in}}{\partial \overline{h}_c} - \frac{\gamma^1}{\mu\pi^1}\frac{\partial V^{1out}}{\partial B^1} p(\overline{q}_c) \\ + \frac{\lambda^{in}}{\mu\pi^1}\frac{\partial \widehat{V}^{2in}}{\partial B^1}\left[\widehat{MRS}^{2in}_{\overline{h}_c B} - p(\overline{q}_c)\right] &- \frac{\lambda^{out}}{\mu\pi^1}\frac{\partial \widehat{V}^{2out}}{\partial B^1} p(\overline{q}_c). \end{aligned} \tag{20}$$

Proof See Appendix A.2. □

Condition (19) has the same structure and interpretation as condition (17). The only difference is that the first term on the right hand side of (19), which accounts for the budget cost incurred by the government when marginally raising $\overline{q}_c$, is simpler than the corresponding term in (17). In particular, given that in the impure opting-out scheme the government directly controls the number of hours $\overline{h}_c$ that are provided free-of-charge to each household who opts in, a marginal increase in $\overline{q}_c$ does not affect the public budget through a behavioral response in h_c (as long as the increase in $\overline{q}_c$ does not induce additional households to opt-in).

Condition (20) characterizes the optimal value for $\overline{h}_c$ by means of a formula which captures the net effect of a policy reform that marginally raise $\overline{h}_c$ while at the same time adjusting B^1 in such a way to leave unchanged the well-being of low-skilled households who opt-in.

The left hand side of (20) captures the impact of such a compensated (for low-skilled households) marginal increase in $\overline{h}_c$ on the externality-term entering the objective function maximized by the government. This term vanishes if the initial equilibrium is such that $h_c^{1in} > \overline{h}_c$; if instead $h_c^{1in} = \overline{h}_c$ at the initial equilibrium, and provided that there are positive externalities associated with an increase in the overall quality of child care arrangement for kids in low-skilled households (i.e. an increase in q^1), the term on the left hand side of (20) will be positive if a (compensated) marginal increase in $\overline{q}_c$ induces an increases in q^1.

On the right hand side of (20) we have instead the effects of a compensated (for low-skilled households) marginal increase in $\overline{h}_c$ on the individual-rationality constraints (the γ-constraints) and the self-selection constraints (the λ-constraints). Starting with the effects on the individual-rationality constraints, a marginal increase in $\overline{h}_c$ has an adverse effect on the γ^2-constraint since it makes more tempting for high-skilled households to opt-in, which represents a cost of expanding $\overline{h}_c$ if the goal of the government is to have only low-skilled households to opt-in. On the other hand the reform will also make more tempting for low-skilled households to opt-in (given that, if they were to opt-out they would only bear the cost of the reform, i.e. the reduction in B^1); provided that the goal of the government is to induce low-skilled households to opt-in, this effect represents a benefit of marginally raising $\overline{h}_c$. Finally, regarding the effects on the self-selection constraints, the proposed reform would lower the utility of high-skilled households who were to behave as mimickers while opting-out (last term on the right hand side of (20)), and at the same time, provided that at the initial equilibrium $h_c^{1in} > \overline{h}_c > \widehat{h}_c^{2in}$, also lower the utility of high-skilled households who were to behave as mimickers while opting-in.

4 Quantitative Model

We now extend the model considered in section 3 and proceed with a quantitative analysis of optimal child care policy.

4.1 General setting

We maintain the assumption that market ability and nurturing ability are positively correlated and that there is assortative mating; this allows us to summarize the ability type of a given household by means of a single parameter. We consider a discrete set of households indexed by their ability type $i \in \{1, \ldots, N\}$, where $N > 2$ and where a higher index corresponds to a higher ability. We also explicitly introduce an extensive participation margin of labor supply for mothers.[16] This is done by assuming that in each household there is a fixed cost associated with the mother's labor force participation. More specifically, we assume that mothers in households of type i differ in their fixed cost type $j \in \{1, \ldots, \Psi\}$ and incur a fixed cost of χ_{ij} when entering the labor force.[17] Thus, the type space is fully characterized by the tuple $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, \Psi\} \equiv \Theta$. This means that, even though we have assumed that household wage rates can

[16]We do not consider an extensive participation margin for fathers. This is without loss of generality as the vast majority of fathers participate in the labor force. The importance of the interaction between mothers' labor force participation decisions and child care costs in the context of a quantitative optimal tax model has previously been emphasized by Blundell and Shephard (2012). The role of child care subsidies for labor force participation and fertility has recently been explored by Bick (2016).

[17]Notice that the model in section 3 already includes the option for mothers to choose to work zero hours. However, the fraction of mothers who would actually choose to remain outside the labor force would be too low in comparison to what we see in the data. Thus, in line with earlier literature (e.g. Cogan 1981, Hausman 1980), we assume that there is a fixed cost associated with the mothers' labor force participation.

be summarized into a unidimensional parameter, the type-space is still bi-dimensional by virtue of the heterogeneity in the fixed costs of work. In section 5 we explain how we deal with this computationally.

The skill level of the household is described by the mothers' relative position (rank) in the wage distribution of mothers.[18] We denote the wage rates of mother and father in household i as (w^i_m, w^i_f). The wage rate of mothers in households of type i is set to the average wage among mothers with skill position i. The wage rate of fathers in household i is equal to the average wage among all fathers who are married to skill type i mothers. The procedure to compute the wages based on actual data is described in detail in section 4.2 below.

4.2 Household decision problem

We perform our calibration of the model under the current US tax system while taking existing child care subsidies into account. The problem solved by each household $(i, j) \in \Theta$ in the presence of existing taxes and child care subsidies, and the extensive margin of labor supply, can be described as follows:

$$\max_{h_m, L_m, L_f, h_c, q_c} \quad v\left(c, h_m, L_m, L_f, h_c, q_c; w^i_m, w^i_f, \omega^i_m, \omega^i_f\right) - \mathbf{1}[L_m > 0] \cdot \chi^{ij} \tag{21}$$

subject to the household budget constraint:

$$c = w^i_f L_f + w^i_m L_m - T^{US}(w^i_f L_f + w^i_m L_m) - CE(p\,(q_c)\,h_c, w^i_f L_f, w^i_m L_m), \tag{22}$$

where T^{US} is the tax function and CE is the *net* (of subsidy) child care expenditure of the household as a function of gross child care expenditure $p\,(q_c)\,h_c$ and the income of both spouses, $w^i_f L_f$ and $w^i_m L_m$. The functions T^{US} and CE are chosen to approximate the rules governing taxes and child care subsidization in the US. In specifying T^{US}, we follow Heathcote et al. (2014) and assume the following parametric form for T^{US}:

$$T^{US}(y) = y - \lambda y^{1-\tau}$$

which implies that the relationship between post-tax income $\widetilde{y}$ and pre-tax income y is given by $\widetilde{y} = \lambda y^{1-\tau}$ or, equivalently, $\log(\widetilde{y}) = \lambda + (1-\tau)\log(y)$ which we estimate by OLS using information on the relationship between $\widetilde{y}$ and y provided by NBER TAXSIM. Using the relevant sample of households filing jointly with small children, and assuming *zero* child care expenditure, we find $\tau = 0.164$ and $\lambda = 1.31$.

The purpose of our quantitative exercises will later be to consider various reforms where we replace the T^{US} function with an optimally chosen nonlinear income tax schedule and where we

[18]In principle, the skill rate of the household could be constructed based on the mothers skill, the fathers skill, or any combination of the two. We have chosen to let the mother dictate the skill level of the household since we focus on their labor force participation decisions.

replace the CE function with different child care subsidy schemes, and compute the associated welfare gains.

4.3 Functional forms

We assume that the utility function of households takes the following form

$$U = \frac{c^{1-\beta}}{1-\beta} - \xi_{1m}\frac{1}{\ell_m} - \xi_{1f}\frac{1}{\ell_f} - \xi_{2m}\frac{1}{h_m} - \xi_{2f}\frac{1}{h_f} + \xi_3\frac{q^\rho}{\rho} - \mathbf{1}[L_m > 0]\cdot\chi^{ij}. \tag{23}$$

In the above specification we have employed functional forms that are standard in the literature and are suitable for numerical computation. The household derives utility from consumption c through a standard CARA term. The utility from leisure time without children, ℓ_m and ℓ_f, as well as the utility from time with children, h_m and h_f, enter through iso-elastic functional forms.[19] The parameters $\xi_{1m}, \xi_{1f}, \xi_{2m}, \xi_{2f}, \xi_3$ are weighting parameters which control the importance of the different time-use components in the utility function.

Notice that, in similarity to how equation (6) was obtained from (5), we always use the time constraints of the mother, father and child to eliminate ℓ_m, ℓ_f, h_f, and obtain the individual decision problem as presented in (21).[20]

We let q, the overall quality of the child care arrangement, be given by

$$q = \kappa^i[\omega^i_m h_m + \omega^i_f h_f + kq_c h_c], \tag{24}$$

which depends on the type-specific parameter κ^i defined as a weighted average of the average market productivity of parents in household i, and the average market productivity in the population, i.e. $\kappa^i = \theta\frac{w^i_m + w^i_f}{2} + (1-\theta)\sum_{j=1}^{N}\frac{w^j_m + w^j_f}{N}$, $\theta \in [0, 1]$. The type-specific parameter κ^i can be thought as capturing the degree of genetic transmission of ability within households; in particular, the larger the value of θ and the larger the degree of genetic transmission of ability within households. The function q can be seen as a production function for the child's human capital. Correspondingly, the term q^ρ/ρ can be interpreted as the utility parents derive from the human capital formation in their child. Finally, the parameter k controls the relative importance of formal vs. maternal/paternal care in the production of human capital.

To construct the functional form for the CE function we model the most important child care subsidies in the US. We consider the two federal tax credits, the CTC (Child Care Tax Credit) and the CDCTC (Child Care and Development Tax Credit).[21] In addition we model the state

[19]The particular functional form for the time components of the utility function has previously been employed by Tuomala, (2010).

[20]Specifically, $\ell_m = 1 - h_m - L_m$, $\ell_f = h_c + h_m - L_f$, and $h_f = 1 - h_m - h_c$. Thus, the variables to be optimized are h_m, L_m, L_f, h_c, and, q_c as in (21).

[21]These are the most important federal tax credits. Additionally, states are offered a block grant from the federal government in the form of the CCDF (Child Care and Development Fund). The purpose of the CCDF is to increase the availability, affordability, and quality of child care services. We describe these rules in more detail in appendix B. Since there is substantial variability across states how these funds are used and as these funds operate mainly

tax credit that applies in California and which is a function of the CDCTC.[22] Since the actual rules governing child care involve various kinks, we calculate the CE function using a smooth approximation to facilitate incorporation in our computational model.[23]

4.4 Calibration

In order to obtain numerical results, values for all exogenous parameters must be specified. This includes all the parameters entering the utility function (23), the human capital production function (24), the market wage rates (w_m^i, w_f^i), the maternal/parental care productivities (ω_m^i, ω_f^i) and the fixed costs χ^{ij} associated with the mothers' labor force participation. The parameters entering the utility function and the human capital production function are presented in table 1.

Table 1: Parameters

Parameter	Value	Parameter	Value
β	0.800	ξ_3	-0.105
ξ_{1m}	0.450	θ	0.750
ξ_{1f}	0.450	k	40.0
ξ_{2m}	0.075	ρ	-0.50
ξ_{2f}	0.075		

The wage rates (w_m^i, w_f^i) are directly estimated from data. The maternal and paternal care productivities (ω_m^i, ω_f^i) are chosen to match empirically relevant patterns of the ratio of market work to household work as well as reasonable patterns for formal care and home care. The fixed costs $\{\chi^{ij}\}_{(i,j)\in\Theta}$ associated with the secondary earner's labor market participation are calibrated to match the empirical skill-specific labor force participation rates. The details of the calibration procedure are described in the following sections.

In our simulations we consider $N = 5$ skill types and a continuous distribution of cost types (approximated in the simulations by $\Psi = 1000$). Computational considerations prevent us from expanding the model beyond five household types. In practice, however, our model features 10 household types, as households where the mother works behave differently than households where the mother is out of the labor force.

through the supply-side of the child care market, we have chosen not to model these additional subsidies as we are considering a fixed supply schedule for child care.

[22] US states offer subsidy schemes which usually are functions of the federal tax credits, which allow states to tailor the subsidies in terms of percentage rates, income thresholds and other eligibility criteria. In principle, states can offer a larger credit for families who enroll children in state-certified facilities.

[23] Note that an alternative would have been to model a tax/transfer schedule while at the same time *including* the child care subsidy schemes. However, since we do not have data on child care expenditure, we have chosen to construct the functions T and CE separately.

Market wage rates, labor market participation, and hours of work Our model economy consists of couples where each spouse has a different market wage rate. To completely represent all the different types of couples in the economy we would need to consider a matrix of household types where each element ij corresponds to the couple where the father has skill type i and the mother has skill type j. Allowing for such a rich type structure would however be computationally intractable. We therefore wish to index households by a unidimensional parameter i, which we refer to as the *household* skill type. Since much of our emphasis will be on the labor market decisions made by mothers, we have chosen to let the skill type of the household correspond to the skill category of the mother. Thus, (w_m^i, w_f^i) refers to the wage-pair in a household where w_m^i represents the female wages in the i:th skill category. Accordingly, w_f^i is the average wage of all fathers matched with type-i mothers.

We use the Current Population Survey (CPS) Labor Extracts 2003-2006 as our main data source. We compute the average wage, labor market participation rate, and work hours for husbands and wives by household skill level (as defined above) for each of the years 2003-2006. In a final step, we average these measures over the years 2003-2006 to make our calculations less sensitive to year specific shocks. Wage rates are obtained by dividing weekly earnings by weekly hours of work. Our sample contains all married couples between age 20 and 65 who were not self-employed, and who had at least one child below the age of 6. All wages are expressed in terms of 2006 USD.

To obtain a wage rate for mothers who lack a wage observation we follow an imputation procedure. We regress log wages on a set of covariates, including flexible controls for age and education. We also include the education and age of the husband in this regression. The regression generates a set of predicted values for mothers who lack a wage observation. However, all these predictions lie on the regression surface. To obtain correct moments of the distribution of female wages, we draw a large number of samples from the empirical distribution of the residuals in the prediction regression and add these to the predicted wages. The final measure of the wage rate for mothers is equal to the actual wage, whenever it exists, and equal to the predicted wage otherwise.[24] The wage distributions for mothers and fathers are approximated using the deciles of the predicted wage distributions.[25]

We focus on hours worked per week, measured in terms of the "usual weekly working hours" during a typical work week. This might be missing some variation that stems from the fact that some workers can have more than one job.[26] In addition, a large part of the variation

[24]This procedure neglects the fact that workers and non-workers might be different along unobservable dimensions, resulting in selection. This is a standard issue in the literature and is usually addressed by adding a selection term to the prediction equation. However, the credibility of such corrections is severely hampered by functional form assumptions and lack of suitable instruments. For robustness, we have performed a selection correction using county as instrument. This turned out to have a very minor impact on the discrete wage distributions that we use in our simulations.

[25]More specifically, we divide the wage distribution into ten bins and then use the median wage rate within each bin to represent each wage category. Note that the CPS data is top-coded but this poses no problem since the median of the top bin always falls below the top-code.

[26]In the CPS, we use the variable "usual hours". There is another variable called "hours worked last week" that

in annual hours of work stems from the number of weeks worked during a year. However, there is no good information on the number of working weeks in the CPS data. There is a variable called "how many weeks a year do you get paid for", but according to the NBER it is reported only by 12 percent of earners, and almost 90 percent of them report working 52 weeks per year. Since we do not want to overfit our model to noisy wage data, our calibration procedure targets average working hours for mothers and fathers. In order to interpret the earnings in the model as annual earnings, we multiply weekly earnings in an ad-hoc fashion by 48.

The spousal wage rates, hours of work, and the labor market participation of mothers associated with the different household types $i = 1, \ldots, 5$ are displayed in table 2.

Table 2: Hourly wage rates (2006 USD), weekly hours of work, and labor force participation rates (LFP) for mothers and fathers .

Type	w_m	w_f	$Hours_m$	$Hours_f$	LFP_m	LFP_f
1	7.00	14.01	31.99	43.03	0.53	0.95
2	10.62	16.07	34.09	43.74	0.61	0.96
3	14.25	18.49	35.61	43.96	0.63	0.96
4	19.40	21.53	35.84	44.40	0.65	0.97
5	30.47	27.39	34.38	44.74	0.72	0.97

Fixed costs The distributions of fixed costs associated with mothers' labor force participation are chosen so that the model, under the benchmark US tax system, exactly matches the household-specific motherly employment rates in table 2.[27] For this purpose, we have proceeded in the following way. For each skill type i we compute the fixed cost that would make an agent of type i indifferent between working and not-working in the calibrated benchmark economy. Denote this fixed cost threshold $\overline{\chi}^i$. Notice that mothers with $\chi^{ij} \leq \overline{\chi}^i$ will work and mothers with $\chi^{ij} > \overline{\chi}^i$ will stay out of the labor force. We further assume that the lower bound of the fixed cost is 0. If the fraction of workers of type i in the data is z^i we want to assign a fixed cost of less than $\overline{\chi}^i$ to a fraction z^i of the workers of type i. This can be achieved by imposing

potentially could capture the labor supply associated with multiple jobs. However, we did not use this variable since it is plagued by measurement error (e.g. some workers report that last week they worked 0 whereas in a *usual* working week they would work 40 hours).

[27]Notice that by setting the fixed cost distributions appropriately, it is always possible to match any particular pattern of empirical participation rates. For example, if the fraction of mothers who work in household of type 3 is 52% and the number of cost types Ψ is equal to 100, we can always set $\chi_{3j} = -\infty, j = 1, \ldots, 52$ and $\chi_{3j} = +\infty, j = 53, \ldots, 100$. However, this would make the labor force participation of type 3 mothers completely inelastic.

that the fixed cost are given by a type-specific linear function:[28]

$$\chi_{ij} = \frac{\overline{\chi}^i}{\Psi z^i - 1}(j-1), \quad j = 1, \ldots, \Psi$$

We make the observations that $\chi^{ij} = 0$ when $j = 0$ and $\chi_{ij} = \overline{\chi}^i$ when $j = \Psi z^i$, where Ψz^i is the index of the highest cost type that is working.

Notice that the slope of the fixed cost distribution is directly related to the concept of "participation elasticity" emphasized in the public finance literature. Our approach implicitly pins down skill-specific participation elasticities through structural assumptions and a calibration procedure.

As evident from table 2, the non-participation of US mothers is highest among low-income (low skilled) households. Mothers belonging to these households will thus, on average, be assigned higher fixed costs through the calibration procedure.

Time endowment As already mentioned, each adult household member is endowed with one unit of time that can be allocated to hours at the job, hours in maternal/paternal care, and leisure. We interpret the unitary time endowment as representing the time available during a year after having deducted the time needed for sleep. Thus, the unitary time endowment corresponds to 5840 hours. Since children aged 0-6 sleep more than adults, the time endowment for the child (i.e. time during which the child's human capital can be affected) is set to 80% of the adult time endowment. Our data set contains hours worked during a usual working week. Assuming each agent works 48 weeks during a year, an agent that works 40 hours per week spends a fraction $\frac{48 \times 40}{5840} \approx 0.33$ of his/her time endowment on the job. Thus, to calibrate, say, the fathers' labor supply to 40 hours a week, we would like to have $L_f = 0.33$. A type 3 father in the model has an hourly wage rate of 14.25 USD and will thus contribute an amount $0.33 * 14.25 \approx 4.70$ to household income. This corresponds to an annual income of $4.70 \times 5840 = 27,448$ USD. Thus, to translate the model output to annual income, the model output should be multiplied by 5840.[29]

Home care productivities We assume that there are no gender-differences in the domestic child care productivities ω_m^i and ω_f^i, and set them both equal to the average market productivity in each household, i.e. $\omega_m^i = \omega_f^i = \frac{w_m^i + w_f^i}{2}, i = 1, \ldots, N$. The four time-use parameters ξ_{1j} and ξ_{2j}, $j = m, f$ are calibrated to ensure empirically relevant time allocation patterns consistent with the American Time Use Survey (ATUS).

[28]Other forms of parametric fixed cost distributions have been employed in the literature. For example, Kleven et al. (2009) use a power distribution.

[29]We fix the number of weeks worked during a year to 48 as the CPS data does not contain a reliable measure of the number of weeks worked during the year. As the total variation in hours of work across skill groups can be decomposed into the variation in hours worked during a typical week, and in terms of the number of weeks worked during a year, we are therefore likely to underestimate the true variation in annual hours of work.

The ATUS data contains measures on the time devoted to various activities during a typical weekday. Table 3 describes how married women with children in ages 0-6 divide their time. To express the time devoted to the various activities as a fraction of the time endowment in our model we proceed in the following way. As already mentioned above, we assume that agents who are employed work 48 weeks per year. Thus, for employed mothers, there are $48 \times 5 = 240$ weekdays and $4 \times 7 + 48 \times 2 = 124$ holidays/weekend days. Correspondingly, using the numbers reported in table 3 we would like to use $\frac{1.9 \times 240}{5840} + \frac{4.2 \times 124}{5840} = 0.1673$ as a target for the fraction of the time endowment employed mothers spend in domestic care. For non-working mothers, the target is $\frac{4.2 \times 52 \times 7}{5840} = 0.2618$. In this calculation we have taken the value for non-employed mother to apply during weekends. To obtain numbers for the domestic child care for men, we compute the ratio between the time male and female respondents report in response to the question "caring for and helping household children" in the ATUS.[30] Thus, we calibrate the model so that fathers in households where the mother works, spend 66% of the time mothers spend in household child care, and so that fathers in households where the mother does not work, spend 40% of the time the mother spends in child care. This yields a fraction of approximately 0.11 in both states. Notice finally, by virtue of the child's time constraint, that center based care is the residual after having subtracted maternal and paternal care. In reality, there are other modes of care available, such as grandparent care. Thus, to account for this fact, and to not overstate the importance of center-based care, we let the levels of maternal and paternal care in our calibration be slightly higher than the number in the ATUS.[31]

The labor supply for a full-time working mother, according to the ATUS, is equal to 6.7 hours per working day, which amounts to $6.7 \times 5 = 33.5$ hours per week. This is roughly in line with the average labor supply reported in the CPS (see table 2). Thus, we calibrate the average labor supply of mothers in the model to approximately be equal to $\frac{33.5 \times 48}{5840} \approx 0.275$. For men, we target an average labor supply of $\frac{44 \times 48}{5840} \approx 0.36$, which is also consistent with the CPS figures reported in table 2.

Price of child care We consider a linear production technology for center-based child care which depends only on labor inputs.[32] More specifically, we assume that the function $p(q_c)$ determining the price of formal child care is equal to $\frac{w_m^3}{8} q_c$. This implies that the quality level $q_c = 1$ corresponds to a child care facility where a mother of skill type 3 takes care of 8 kids.

4.5 Allocations in the calibrated economy

In table 4 we show the allocation for the benchmark calibrated economy where households face the current US tax system and existing child care subsidies. In table 5 we give a more

[30] We use tables 8B and 8C at https://www.bls.gov/news.release/atus.toc.htm.

[31] We have considered including grand-parent care in our model, however it would further enhance the computational challenges as it would add additional decision-variables in the household decision problem.

[32] This assumption is made for simplicity. In reality there are other aspects of quality such as the location of the child care center or the quality of the food served in the child care center.

Table 3: Weekday time use of married women living with young children, by employment status (average hours per day)

	Not employed	Employed part-time	Employed full-time
Sleeping	8.5	8.5	8.2
Household activities	3.8	2.2	1.6
Caring for household children	4.2	2.7	1.9
Working and related activities	≈ 0	3.7	6.7
Leisure and sports	3.5	3.3	2.4

Note: Data include all married women, ages 25 to 54, with a child under 6 present in the household. Data include non-holiday weekdays and are annual averages for 2015.
Source: Bureau of Labor Statistics, American Time Use Survey, 2015
(https://www.bls.gov/tus/charts/chart2.txt)

detailed overview of the patterns of time use in our calibrated model. The top panel in this table describes the time allocation for mothers and fathers in families where the mother works, and the bottom panel describes the time allocation for parents in families where the mother does not work.

Table 4: Benchmark allocation (calibrated economy)

Allocation in households where the mother works

Household	y	c	CE/D	D/y	T/y	g	q	h_c	U
1	37.27	35.365	0.414	0.213	0.034	35.145	2.086	0.367	1.308
2	48.29	42.692	0.523	0.184	0.075	39.624	2.211	0.386	2.014
3	60.63	50.520	0.634	0.160	0.109	43.863	2.362	0.394	2.731
4	78.00	60.905	0.771	0.139	0.146	49,929	2.605	0.401	3.704
5	117.06	82.530	0.989	0.117	0.202	64.069	3.219	0.410	5.714

Allocation in households where the mother does not work

Household	y	c	CE/D	D/y	T/y	g	q	h_c	U
1	35.51	33.350	1.000	0.109	0.026	19.889	1.133	0.330	1.135
2	41.40	37.092	1.000	0.117	0.051	24.959	1.419	0.329	1.771
3	48.36	41.496	1.000	0.122	0.075	30.353	1.722	0.329	2.431
4	57.20	46.793	1.000	0.128	0.101	37.777	2.144	0.328	3.327
5	74.61	56.529	1.000	0.138	0.140	53.422	3.047	0.326	5.189

Income y and consumption c expressed in thousands of USD (2006 values).

Notation: L_j = labor supply, h_j=home care ($j = m, f$), h_c = paid care, y=total family income, c= family consumption, CE=net child care expenditure, D=gross child care expenditure, T= income tax liability, q=overall quality of child care arrangements, q_c = quality level of paid care arrangement. Notice that $CE/D = 1$ for all households where the mother does not work since such families are ineligible for child care subsidies in the current US tax system.

Table 5: Household time allocation (calibrated economy)

Mother time allocation (families with working mothers)

Household	time with child (h_m)	time w/o child (ℓ_m)	labor supply (L_m)
1	0.255	0.612	0.133
2	0.229	0.549	0.221
3	0.217	0.517	0.266
4	0.205	0.487	0.308
5	0.189	0.447	0.363

Father time allocation (families with working mothers)

Household	time with child (h_f)	time w/o child (ℓ_f)	labor supply (L_f)
1	0.178	0.432	0.389
2	0.185	0.447	0.368
3	0.189	0.454	0.357
4	0.194	0.463	0.343
5	0.201	0.472	0.328

Mother time allocation (families with non-working mothers)

Household	time with child (h_m)	time w/o child (ℓ_m)	labor supply (L_m)
1	0.304	0.696	0
2	0.306	0.694	0
3	0.308	0.692	0
4	0.311	0.689	0
5	0.316	0.684	0

Father time allocation (families with non-working mothers)

Household	time with child (h_f)	time w/o child (ℓ_f)	labor supply (L_f)
1	0.166	0.399	0.434
2	0.165	0.394	0.441
3	0.163	0.389	0.448
4	0.161	0.384	0.455
5	0.158	0.375	0.466

The column CE/D in table 4 shows the fraction of child care expenditure that the household pays for itself. Since households where one spouse does not work are ineligible for the subsidies that we consider in the calibrated model, this fraction is equal to one for all one-earner households. For two-earner households the effective subsidy ranges between 58.6% and 1%, and is monotonically decreasing in the skill type of an household.

The column T/y reports the average income tax rate paid by the various households, which ranges between 2.6% (for one-earner households of type 1) and 20.2% (for two-earner households of type 5). The column q shows that the overall quality of the child care arrangement is increasing in the skill type of an household and, for each given skill type, is higher in two-earner households than in one-earner households. Moreover, one can also see that the gap between the overall quality of child care arrangement chosen by two-earner households and one-earner households shrinks as the considered household skill type increases. Finally, the column h_c shows that among two-earner households there is more variation in the number of hours of center-based care chosen by parents.

5 Optimal Tax Systems

5.1 Government objective function

To capture the welfare gains of policy reform we assume the government maximizes a social welfare function of the Utilitarian type, aggregating individual utilities into a scalar measure of social welfare, putting equal weight on the well-being of each individual. In addition, the social welfare function captures externalities relating to the overall quality of child care that children in the economy are exposed to. Formally, we add the following expression to the social welfare function:

$$v(\mathbf{q}) = \overline{v} \sum_{i} \left(\overline{q} - \frac{a_1}{(q^i)^{a_2}} \right), \tag{25}$$

where we set $a_1 = 1$, $a_2 = 2$, and $\overline{v} = 60$. The above term captures how the human capital formation in the child generation impacts the welfare of the adult population. The term is large and negative when $q^i, i = 1, \ldots, N$ takes on small values and the contribution to the externality by children of type i converges to $\overline{q}$ when q^i is large. In other words, agents' care about the level of human capital in other parents' children, but at a rate which is declining and reaches zero at "acceptable" levels of human capital. This captures the common view that all children should have access to child care of adequate quality and that no children should be left behind. The parameters a_1 and a_2 are location and curvature parameters associated with the functional form for the externality components. The weight $\overline{v}$ controls the overall strength (importance) of the above considerations for social welfare. When $\overline{v} \neq 0$, welfare gains from policy reform will reflect both the extent to which the policy instruments mitigate the distortions associated with income taxation, and to which extent the policy instruments cause parents to make better choices for their children, captured by the external effects of children's human capital on society.

5.2 Computational approach

The problem of finding the optimal tax and child care policy represents a bi-level programming problem. To evaluate the social welfare level associated with a particular policy set by the government it is necessary to compute how agents optimally respond to this policy. Thus, there is an upper level (government) optimization problem and a lower level optimization problem that is solved by each type of household in the economy.

The challenges associated with solving bi-level optimization problems numerically are well-known. The difficulties usually derive from the need to impose the first-order conditions to the agents' problem as nonlinear equality constraints in the government's optimization problem.[33] Given the large number of private decision variables, we did not find a procedure that incorporates the first-order conditions as constraints to be very robust. Instead, we compute the solutions to the individual decision problems numerically using a nested optimization procedure. In contrast to the first-order approach, this procedure allows us to take into account both first and second order conditions in the individual optimization problem. The drawback is that we have to rely on numerical approximations of derivatives in the upper level which significantly increases the time it takes to find an optimal solution. In addition there is a computational overhead associated with the nested optimization layer. To increase performance, exact first and second order derivatives to the lower level optimization problem were provided to the numerical optimization algorithm and we relied on a fast and efficient implementation using AMPL and C++.

The presence of an extensive margin of labor supply for mothers and the heterogeneity in the fixed cost of working imposes particular challenges for finding the solution to the government's problem. Perhaps most fundamentally, since we have both heterogeneity in the fixed costs of working and in skills, the government's problem is a multidimensional screening problem. Such problems are inherently complex to solve since designing a fully nonlinear income tax implies that the government screens workers by offering a distinct contract to each type of agent subject to a set of self-selection constraints. When the type space is multi-dimensional, unless the number of types in each dimension is very small, achieving an incentive-compatible allocation requires that a very large number of incentive constraints be satisfied.[34]

To achieve tractability, and reduce the type-space, we assume that the fixed cost of work is a utility cost entering additively in the utility function. This implies that, among equally skilled households, all households will make the same choices regarding the individual decision variables provided that the mother has the same labor force participation status and the same skill level. Moreover, we know that among equally skilled mothers, those with a higher fixed cost will always be less likely to participate in the labor force.[35] This allows us to identify, for each skill group, a unique marginal worker that is indifferent between working and not-

[33] Similar challenges appear in dynamic mechanism design problems where savings are assumed to be unobservable to the social planner. In our setting, after all possible substitutions have been made, there are not one but four privately chosen variables that are handled in the subproblem. These are: the labor supply of the mother, the hours of maternal care, hours of formal child care, and the quality in formal care.

[34] For a discussion about the exponential increase in the number of self-selection constraints in a multidimensional screening setting, see Bastani et al. (2013). In the present case, due to the complexity of the individual subproblem, each additional incentive constraint that needs to checked entails a substantial computational cost.

[35] In contrast, if the fixed cost of work was modeled as a monetary cost, there would be a countervailing income effect.

working. Mothers with a fixed cost greater than the marginal worker will always stay out of the labor force, and mothers with a lower fixed cost than the marginal worker, will be working. This means that at each skill level, we only need to compute the optimal individual decisions for a representative two-earner household and for a representative one-earner household, rather than computing these decisions for each possible fixed cost type.[36] It also implies that the government only needs to design two set of bundles for each type i. One pre-tax/post-tax income point for two-earner households of type i and one pre-tax/post-tax income point for one-earner households of type i. This drastically reduces the number of incentive constraints that need to be incorporated into the government's problem, and also allows us to employ a large number of discrete cost types.

The above simplifications notwithstanding, there are three main obstacles towards increasing the number of skill types that we consider. First, for every additional type one needs to compute additional individually optimal decisions (i.e. hours of work and child care decisions), which requires additional computational resources. Second, for every additional agent we introduce in the economy, we need to expand the set of pre-tax/post-tax income points offered by the government, which increases the number of control variables that need to be optimized in the "main" government problem. These additional income points also generate additional self-selection constraints, making it more difficult to achieve convergence in the main problem. Finally, and perhaps most critically, as explained below, adding types increases the number of marginal workers that need to be identified in order to determine the number of mothers who find it optimal to work.

There are two approaches to modeling the extensive margin. One approach is to let agents optimally choose their labor force participation status in the lower level optimization problem. This implies that the fraction of workers at each skill level is endogenous to the tax system. While this does not introduce any non-smoothness in the government's social welfare function or tax revenue function (provided the number of cost types is sufficiently large), it does imply that individuals might switch discretely from working to not working, or vice versa, in response to a small change in the income tax. This causes an undesirable reshaping of the set of incentive constraints, which makes it difficult to find solutions to the government's problem using gradient-based optimization algorithms. We have therefore refrained from this approach. Instead, we add the binary variables associated with mothers' labor force participation decision as artificial control variables of the government, while adding a set of constraints ensuring that the labor force participation decisions assigned to agents are incentive-compatible. The benefit of this approach is that the marginal control variables can be treated as exogenous and optimized in a separate optimization layer. This means that our optimization problem has three layers. An outer layer where we choose the labor force participation levels at each skill level (equivalent to identifying the marginal worker), a middle layer where we choose the pre-tax/post-tax income

[36] Notice that without the assumption that the utility cost is additive, there would be an explosion in the number of individual decision problems that need to be computed, making the problem computationally intractable.

points as well as the child care subsidy instruments, and a bottom layer, where agents make optimal decisions taking the tax policy environment as given.[37] For the upper layer, as will be explained in more detail below, we rely on a customized global search of the parameter space which has a computational complexity similar to a grid search. We therefore employ all our parallel computing resources at the upper level.[38]

The labor force participation decision of mothers is represented by a binary matrix $\mathbf{L}$ where $\mathbf{L}_{ij} = 1$ if the mother of type (i, j) is working, and zero otherwise. Since the fixed cost of work χ_{ij} is assumed to be non-decreasing in j, the rows of $\mathbf{L}$ will be non-increasing when moving from the left to the right. This allow us to introduce the vector $\mathbf{P}$ where $\mathbf{P}_i$ is the number of leading ones along row i. For example, suppose $N = 5$ and $\Psi = 10$ and that

$$\mathbf{L} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

In this case, by counting the leading ones along each row, we have $\mathbf{P} = \begin{bmatrix} 2 & 3 & 5 & 8 & 9 \end{bmatrix}$. Thus, the vector $\mathbf{P}$ describes the participation rate at each skill level and the participation rates are given by $\mathbf{P}/\Psi = \begin{bmatrix} 0.2 & 0.3 & 0.5 & 0.8 & 0.9 \end{bmatrix}$. Notice that $\mathbf{P}_i$ is also equal to the fixed cost type of the worker who is, at the margin, indifferent between working and not working. The fixed cost of the marginal worker among households of type i can be computed as $x^i = F_\chi^{-1}(\mathbf{P}_i)$ if $F_\chi(x) \in [0, 1]$ is the CDF of the fixed cost distribution.

5.3 The government's problem

We are now in a position to formally present the problem solved by the government. Assume that the government maximizes a Utilitarian social welfare function and that two distinct non-linear income tax schedules apply to one-earner and two-earner households.[39] In the absence of any kind of subsidies to child care expenditures (i.e. in the case of a pure income tax optimum),

[37] In the analysis of opting out public provision schemes, we also optimize over individuals' discrete binary decision to opt in or opt out of the publicly provided child care services.

[38] The model was solved on a dual processor Intel Xeon workstation with 20 computational cores and 128GB of RAM.

[39] This only requires that the labor force participation decision is observable by the government.

the government's problem can be described as follows:

$$\max_{\mathbf{P}} \Omega(\mathbf{P}) \tag{26}$$

$$\Omega(\mathbf{P}) = \max_{\{(M_1^i,B_1^i),(M_0^i,B_0^i)\}_{i=1}^N} \sum_{i=1}^{N} \left(\sum_{j>\mathbf{P}_i} \pi^{ij} V_0^i(M_0^i, B_0^i) + \sum_{j\leq \mathbf{P}_i} \pi^{ij} \left(V_1^i(M_1^i, B_1^i) - \chi_{ij} \right) \right) \tag{27}$$

$$\text{subject to:} \tag{28}$$

$$V^{ij}\left(M_0^i, B_1^i, M_1^i, B_1^i\right) \geq \tilde{V}^{ij}\left(M_0^{i-1}, B_1^{i-1}, M_1^{i-1}, B_1^{i-1}\right), \forall i \in \{2, \ldots, N\}, \forall j \tag{29}$$

$$V^{ij}\left(M_0^i, B_1^i, M_1^i, B_1^i\right) = \begin{cases} V_0^i(M_0^i, B_0^i) & \text{if } j > \mathbf{P}_i \\ V_1^i(M_1^i, B_1^i) - \chi^{ij} & \text{if } j \leq \mathbf{P}_i \end{cases} \tag{30}$$

$$\tilde{V}^{ij}(M_0^{i-1}, B_1^{i-1}, M_1^{i-1}, B_1^{i-1}) = \max\left\{V_0^i(M_0^{i-1}, B_0^{i-1}), V_1^i(M_1^{i-1}, B_1^{i-1}) - \chi^{ij}\right\} \tag{31}$$

$$V_0^i(M_0^i, B_0^i) > V_1^i(M_1^i, B_1^i) - \chi^{ij} \quad \forall i, \quad j > \mathbf{P}_i \tag{32}$$

$$V_0^i(M_0^i, B_0^i) < V_1^i(M_1^i, B_1^i) - \chi^{ij} \quad \forall i, \quad j \leq \mathbf{P}_i \tag{33}$$

$$\sum_i \left(\sum_{j\leq \mathbf{P}_i} \pi^{ij}(M_0^i - B_0^i) + \sum_{j>\mathbf{P}_i} \pi^{ij}(M_1^i - B_1^i) \right) \geq R \tag{34}$$

$$V_0^i(M, B) = \max_{q_c,h_c,h_m} u^i\left(B - p(q_c)h_c, h_m, 0, \frac{M}{w_f^i}, h_c, q_c\right), \forall i \tag{35}$$

$$V_1^i(M, B) = \max_{q_c,h_c,h_m,L_m} u^i\left(B - p(q_c)h_c, h_m, L_m, \frac{M - w_m^i L_m}{w_f^i}, h_c, q_c\right), \forall i \tag{36}$$

The first thing to notice is that, as discussed in section 5.2, the government's problem features three levels of optimization. Equation (26) defines the upper level optimization in which the government chooses the participation rate at each skill level to maximize $\Omega(\mathbf{P})$. The function $\Omega(\mathbf{P})$ is in turn the value function associated with the middle or "main" layer of optimization where the government strives to find the income tax schedule (defined in terms of the pre-tax/post-tax income points) that maximizes a social welfare function. The social welfare function is of the Utilitarian type and is equal to the sum of the indirect utilities of all one-earner households and the indirect utilities of all two-earner households, weighted by the population shares π^{ij}. Notice that in the main optimization problem, the parameters $\mathbf{P}_i$ are treated as exogenous.

Turning now to the constraints of the main optimization problem, the set of incentive-constraints appear in (29). These constraints ensure that each household prefers the bundle assigned to it rather than the bundle intended for the adjacent lower skilled household. Equations (30) and (31) define the left hand side and right hand side of the incentive constraints where the parameters $\{\mathbf{P}_i\}_{i=1}^N$ determine whether the relevant utility for an agent of type (i, j) is that which arises if the mother is not working (V_0^i) or that which arises if the mother is working ($V_1^i - \chi^{ij}$). Notice that equation (31) implies that if a type i household decides to mimic a household of type

$i-1$, it must replicate the labor force participation decision of type $i-1$.[40] Inequalities (32) and (33) are individual rationality constraints that ensure that the labor force participation decisions prescribed in the **P** vector are actually the ones maximizing the household utility. Constraint (34) is the government budget constraint stating that the sum of tax revenue from one-earner and two-earner households should sum up to the exogenous revenue requirement R.[41] The last two equations define the indirect utilities in the case where the mother does not work (eq. 35) and the indirect utilities (gross of the fixed cost of work) in the case where the mother works (eq. 36). The computation of these two indirect utilities for each type-i household represents the lower level optimization problem.[42]

For the upper layer, that is responsible for finding the optimal participation vector **P**, we use a global optimization heuristic that relies on a combination of local searches around the calibrated labor force participation rates and coarse searches over the full parameter space.[43] For the middle and lower layers, i.e. the bi-level optimization problem, we rely on an efficient implementation in C++ which interfaces the latest version of the state-of-the-art solver for nonlinear constrained optimization problems KNITRO.

5.4 Welfare Gains Measure

In order to evaluate the welfare gain associated with a given policy reform, we compute the minimum amount of extra revenue that needs to be injected into the benchmark economy, equally distributing it across all households in the economy, in order to reach the post-reform social welfare level (allowing households to re-optimize). We then divide this amount of extra revenue by the aggregate pre-tax income in the benchmark case to get a welfare measure expressed as a fraction of aggregate output.[44]

[40] This is a weak simplifying assumption. The assumption that it is only possible to mimic adjacent types is potentially stronger, as letting the mother drop out of the labor force could be a way for a high skill household to replicate the taxable income of a much more low-skilled two-earner household.

[41] In our calculations below, the revenue requirement R is always set equal to the fiscal surplus that arises in our US benchmark economy.

[42] Notice that these utilities must be evaluated both when a household acts truthfully and when the household behaves as a mimicker.

[43] For computational tractability, we limit the precision of the search to steps of five percentage points in each dimension of P. In addition, we impose that the labor force participation is monotonically increasing in the household skill level, i.e. $P_i \geq P_j$ for all $i \geq j$, $i, j \in \{1, \ldots, N\}$. Finally, we impose that the maximum employment rate at any skill level is 95%, reflecting the realistic assumption that there is a non-negligible fraction of the population with very high fixed costs of work, who would not be willing or able to work regardless of the financial incentives.

[44] We have chosen this approach because it is simple and transparent. There are many potential ways that the revenue that we inject into the benchmark economy can be distributed to agents. One alternative way would be to rebate back the revenue through a linear progressive income tax. Another approach would be to let the government use the injected tax revenue to finance an EITC.

6 Quantitative Results

We now turn to our main results. We will start by analyzing the case where the only tax instrument at the government's disposal is a nonlinear income tax, which we refer to as case (i). Then we will expand the armoury of policy instruments by considering ii) a set of deductibility rates that only depend on the mother's employment status, (iii) a set of tax credit rates that only depend on the mother's employment status, (iv) a set of tax deductibility rates that depend on both the mother's employment status and the household income, (v) a set of tax credit rates that depend on both the mother's employment status and the household income, (vi) an opting-out public provision scheme where the quality of the publicly provided care is chosen by the government and the rate at which, for opting-in households, child care expenses are subsidized is allowed to depend on the mother's employment status, and finally (vii) an "impure" opting-out public provision scheme where the quality of the publicly provided care is chosen by the government and the number of hours that are provided free of charge is allowed to depend on the mother's employment status.

As already explained in section 5.1, to evaluate the welfare effects of policy reform we rely on a social welfare function aggregating household utility levels into a scalar measure of social welfare. This social welfare function also includes an additional human capital term capturing externalities associated with the human capital formation in children. The externality term might represent, for example, the support for the principle of equality of opportunity or a concern that children who are exposed to child care of poor quality will be a burden to society in the future. The latter might result in the need to increase future tax revenue.[45] We envision that the externalities we have in mind are primarily a function of the *distribution* of human capital of the children in the economy. In particular, it is socially desirable to avoid very low levels of human capital.

Our first result concerns the desirability of subsidizing child care expenses in the case when no weight is attached to the externality term, i.e. $\bar{\nu} = 0$ in (25). This implies that the tax instruments are only used for the purpose of relaxing the self-selection constraints faced by the government in the design of the optimal nonlinear income tax. In this case we find that there is no role for using tax deductions or tax credits. The result applies both to the case when the fraction of child care expenditures that can be deducted against the household earned income, or that can be credited against the gross income tax, only depends on the mother's employment status, and the case when the fraction is allowed to depend both on the mother's employment status and the household income. Rather than subsidizing child care expenses, what would be required to mitigate the binding self-selection constraints is to levy a *tax* on the purchase of child care services.

This is an interesting finding as it challenges a standard result in the optimal tax litera-

[45] Thus, increased human capital in children benefits not only children, but also parents (through the household utility function). This means that we are essentially double-counting, as in Kaplow (2009) and Farhi and Werning (2010).

ture that goods complementary to work should be subsidized as a way to relax the incentive constraints associated with income redistribution. The result has its roots in the Atkinson and Stiglitz (1976) theorem on the usefulness of commodity taxes in the presence of a general (nonlinear) labor income tax. According to that theorem, if the income tax is allowed to be nonlinear, commodity taxes are a redundant policy instrument when preferences are separable between leisure and other goods. Instead, if the separability condition is not satisfied, one should use commodity taxes and subsidies to discourage the consumption of goods/services that are substitutes with labor supply and encourage the consumption of goods/services that are complements with labor supply. Given that child care services are regarded as a primary example of goods that are complements with labor supply, a consequence of the Atkinson and Stiglitz (1976) theorem is that they should be subsidized or in any case be subject to a more lenient tax treatment (compared with other goods).[46]

The reason why we get an opposite result is due to the fact that, while previous contributions assumed the hourly price of child care as fixed, we allow child care expenditures to depend both on the number of hours spent by a kid at a child care center and on the quality of the facility chosen by parents (which affects the hourly price of child care services). Thus, while in a model with a fixed hourly price of child care services a low-skilled agent spends more on child care services than a high-skilled mimicker (since a high-skilled mimicker needs to work fewer hours than a low-skilled agent, and therefore needs fewer hours of child care for the kids), this is no longer necessarily true in our setting where the quality (and therefore the hourly price) of child care services is a choice variable for households. In fact, choosing the model's parameters in order to obtain a realistic calibration, we find that a household behaving as a mimicker (i.e. choosing the income point intended for a lower skill type) would spend more on child care than the household being mimicked, and this despite the fact that a mimicker demands fewer hours of center-based child care. The result is due to the fact that the quality, and therefore the hourly price, of child care services chosen by a mimicker is higher than the one chosen by the lower-skilled household being mimicked.

While we find no mimicking-deterring role for subsidizing child care expenses through tax deductions or tax credits, we find that supplementing an optimal nonlinear income tax with an opting-out public provision scheme is welfare-enhancing even when no weight is attached to the externality term.[47] As one can see from table 14 and table 15, the welfare gains amount to 0.08%

[46]This is also the view expressed by Crawford et al. (2010) in one of the chapters contained in the Mirrlees Review (2010).

[47]In the optimal tax literature there are previous examples of contributions where public provision of child care services has been found to be desirable as a device to mitigate binding self-selection constraints (see, e.g., Cremer and Gahvari, 1997) or, more generally, as a device to counteract the distortions induced by income taxation (see, e.g., Gahvari, 1994, 1995). However, these earlier contributions considered a uniform topping-up provision scheme rather than an opting-out scheme. Under a uniform topping-up provision scheme, every agent gets a uniform amount of the good/service that is publicly provided and the publicly provided amount can be topped up with private purchases in the market. In our setting where child care expenditures depend both on the quality of the chosen facility and the number of hours spent by a kid at a child care center, a uniform topping up public provision scheme would require giving all households a voucher that can be used to finance part of their child care expenses.

when the government provides at a subsidized rate child care services of a given quality,[48] and is slightly larger, being equal to 0.11%, for the case of an "impure" opting-out scheme.[49] At an optimum, for both opting-out schemes the quality of the publicly provided care is set at a level that is in between the one chosen, at a pure income tax optimum, by households of type 1 and households of type 2.

For the first opting-out scheme (see table 14) the rate at which child care purchases are subsidized for opting-in households is equal to 14.6% for two-earner households and to 40.4% for one-earner households. This implies that only the first two types opt-in among two-earner households whereas all one-earner households, with the exception of those at the top of the skill distribution (type 5), find it optimal to opt-in.

For the "impure" opting-out scheme (see table 15) the number of child care hours that are provided free of charge to households who opt-in is significantly larger for one-earner households than for two-earner households (0.35 for the former compared to 0.06 for the latter). As a consequence, also in this case only the first two types opt-in among two-earner households whereas all one-earner households, with the exception of those at the top of the skill distribution (type 5), find it optimal to opt-in.

Thus, the results obtained from the analysis of the two alternative opting-out provision schemes are consistent in indicating that, viewed purely as a mean to relax the binding self-selection constraints, public provision should be primarily targeted towards one-earner households under a utilitarian social welfare function. The intuition is that the marginal utility of consumption tends to be higher among one-earner households than among two-earner households, and is therefore towards the former group that redistribution is primarily geared.

Our discussion of the results stated in Propositions 1 and 2 helps understanding why tax deductions or tax credits do not work as mimicking-deterring devices whereas an opting-out provision scheme does. The reason is that for the case of tax deductions or tax credits, what is crucial in assessing their effect on the self-selection constraints is the difference between the amount spent on child care by a true low-skilled household and the amount that would be spent by a high-skilled behaving as a mimicker. In our setting a high-skilled mimicker would spend more than a true low-skilled household (because of the higher quality of the center-based care that it chooses), and this implies that tax deductions or tax credits would be more valuable for a mimicker than for a true low-skilled, undermining the possibility to use these instruments for mimicking-deterring purposes. In the case of an opting-out public provision scheme, instead,

Such a provision scheme would not be welfare-enhancing in our setting for the same reason why tax deductions or tax credits do not work, namely the fact that, when behaving as a mimicker, a high-skilled household spends more on child care services than the low-skilled household being mimicked.

[48] This is the opting-out provision scheme that we discussed in Section 3.1.2. The value of 0.08% for the welfare gain is calculated as the difference between the welfare gains reported in table 14 and the welfare gains reported in table 17 in appendix C: 4.63% – 4.55%.

[49] This is the opting-out scheme that we discussed in Section 3.1.3. The value of 0.11% for the welfare gain is in this case calculated as the difference between the welfare gains reported in table 15 and the welfare gains reported in table 17 in appendix C: 4.66% – 4.55%.

given that the quality of the publicly provided care is set by the government and is the same for all households who opt-in, a subsidy yields a smaller benefit to a low-skilled household than to a mimicker only if the mimicker opts-in and at the same time demands more hours of center-based care than a true low-skilled (which is unlikely to happen given that a high-skilled mimicker needs to work fewer hours than a low-skilled).

While we believe that the result that tax deductions or tax credits are not warranted for mimicking-deterring purposes is interesting from the perspective of the optimal income tax literature, another potential motivation for child care subsidies is that child care of good quality benefits children. Thus, hereafter we will focus on the results obtained when the weight on the externality term is non-zero ($\overline{v} \neq 0$). In this case, the tax instruments are used not only for the purpose of mitigating self-selection constraints, but also for the purpose of affecting the distribution of human capital in the economy. The resulting mothers' labor force participation under the various tax regimes that we consider is displayed in table 6.

Table 6: Mothers' labor force participation in the different regimes

Policy/Household	1	2	3	4	5
Optimal Tax	45%	60%	70%	80%	95%
Deduction	35%	55%	70%	80%	95%
Tax Credit	35%	55%	70%	80%	95%
Deduction (means-tested)	40%	55%	70%	80%	95%
Tax Credit (means-tested)	40%	55%	70%	80%	95%
Public Provision I	40%	55%	70%	80%	95%
Public Provision II	40%	55%	70%	80%	95%

6.1 Pure Income Tax Optimum

We first analyze the case when the government optimizes a nonlinear income tax and there are no subsidies to child care. The results are shown in table 7

Table 7: Pure Nonlinear Income Tax Optimum

Allocation in households where the mother works

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$
44.757	43.965	0.208	0.443	0.235	0.163	0.402	1.174	23.047	0.11	-0.043
55.69	51.058	0.275	0.412	0.215	0.173	0.412	1.482	29.597	0.114	0.078
69.859	59.9	0.318	0.402	0.202	0.176	0.422	1.821	36.917	0.114	0.119
90.436	71.826	0.361	0.394	0.189	0.179	0.432	2.311	47.477	0.115	0.147
149.113	104.25	0.456	0.425	0.16	0.169	0.47	3.429	74.043	0.113	-0.004

Allocation in households where the mother does not work

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$
37.168	35.164	0	0.454	0.303	0.16	0.336	1.158	20.459	0.109	0.09
41.012	37.091	0	0.437	0.306	0.166	0.328	1.423	24.959	0.118	0.22
46.738	39.848	0	0.433	0.309	0.168	0.323	1.694	29.704	0.122	0.291
54.339	43.016	0	0.432	0.313	0.169	0.319	2.06	36.119	0.126	0.352
86.972	59.151	0	0.544	0.315	0.135	0.351	2.975	54.73	0.125	-0.012

Welfare Gain = 3.94% of GDP

Rows correspond to household types 1-5. Taxable income M and consumption c expressed in thousands of USD (2006 values).

As one can see comparing the row "Optimal Tax" of table 6 with the column "LFP_m" of table 2, replacing T^{US} with the optimal nonlinear income tax generates an increase in the proportion of one-earner households among the lowest skilled group (i.e. type 1): the labor force participation of mothers in this skill group drops from 53proportion of one-earner households is left virtually unaffected. For the remaining types, instead, one observes an increase in the labor force participation of mothers, with the increase being more pronounced as one considers more skilled households.

Regarding the optimal profile of marginal income tax rates, one can see from the last column of table 7 that the optimal income tax schedule yields an inverted-U shaped pattern of marginal tax rates inducing more dispersion in labor supply across household types than under the current US tax schedule.

Finally, comparing the column q in table 4 (the benchmark allocation) and 7, one can notice that while the optimal nonlinear income tax does not affect significantly the distribution of the overall quality of child care arrangement among one-earner households, it induces more variation in the corresponding distribution for two-earner households. Interestingly, this result is obtained even though for two-earner households the optimal income tax raises the utility for all but the top skilled households.

6.2 Tax Deductions

We turn next to the case where the government optimizes a nonlinear income tax and subsidizes child care expenses via tax deductions. Assuming that the deductibility rate can be conditioned on the mother's employment status, we separately consider the case when the deductibility rate is uniform within each of the two groups (results displayed in table 8) and the case when it is allowed to be income-dependent (results displayed in table 9).

Table 8: Uniform Deduction

Allocation in households where the mother works

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Ded
44.343	43.168	0.204	0.44	0.236	0.164	0.399	1.165	22.823	0.109	-0.017	0
57.029	51.571	0.284	0.42	0.212	0.171	0.417	1.476	29.756	0.112	0.046	0
70.246	59.927	0.32	0.404	0.201	0.176	0.423	1.817	36.926	0.114	0.113	0
91.38	72.359	0.365	0.398	0.188	0.178	0.435	2.312	47.664	0.114	0.131	0
149.011	104.389	0.455	0.425	0.16	0.17	0.47	3.433	74.095	0.113	-0.004	0

Allocation in households where the mother does not work

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Ded
31.736	33.018	0	0.421	0.302	0.17	0.328	1.293	21.92	0.139	0.231	61.8%
35.534	34.949	0	0.416	0.304	0.172	0.324	1.66	27.872	0.158	0.311	61.8%
41.496	37.943	0	0.423	0.305	0.17	0.324	2.013	33.915	0.164	0.343	61.8%
47.286	40.327	0	0.418	0.307	0.172	0.321	2.554	42.556	0.18	0.416	61.8%
79.674	58.725	0	0.54	0.315	0.136	0.35	2.984	54.744	0.136	0.01	61.8%

Welfare Gain = 4.01% of GDP

Rows correspond to household types 1-5. Taxable income M and consumption c expressed in thousands of USD (2006 values).

Table 9: Means-Tested Deduction

Allocation in households where the mother works

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Ded
35.882	40.239	0.169	0.416	0.247	0.171	0.382	1.275	23.872	0.141	0.116	100%
48.402	49.464	0.271	0.409	0.216	0.174	0.411	1.599	31.469	0.141	0.111	100%
68.456	59.462	0.317	0.402	0.202	0.176	0.422	1.844	37.3	0.118	0.125	16.9%
90.626	71.71	0.362	0.395	0.189	0.179	0.433	2.307	47.436	0.115	0.146	0%
131.974	103.607	0.455	0.424	0.161	0.17	0.47	3.433	74.048	0.127	0.005	100%

Allocation in households where the mother does not work

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Ded
27.914	31.449	0	0.403	0.3	0.175	0.325	1.505	24.556	0.182	0.304	100%
34.12	34.472	0	0.413	0.303	0.173	0.325	1.757	29.153	0.174	0.325	77.8%
44.161	38.851	0	0.429	0.307	0.169	0.324	1.842	31.67	0.141	0.316	34.8%
54.533	42.957	0	0.434	0.313	0.168	0.319	2.056	36.092	0.125	0.349	0%
75.311	58.17	0	0.539	0.315	0.136	0.349	3.005	55.032	0.145	0.021	100%

Welfare Gain = 4.19% of GDP

Rows correspond to household types 1-5. Taxable income M and consumption c expressed in thousands of USD (2006 values).

Table 8 shows that the optimal level of the uniform deduction is 61.8% for one-earner households while no deductibility should be granted for child care expenditures incurred by two-earner households. The possibility for single-earner households to deduct part of the child care expenses is accompanied by an increase in the marginal tax rate that is especially pronounced for the two types at the bottom of the skill distribution. Despite this, among one-earner families the marginal tax rates are increasing for the first four skill types and this implies that the effective subsidy (which is given by the product of the marginal tax rate and the deductibility rate) reaches its maximum ($0.618 \times 0.416 = 25.7\%$) for one-earner households of type 4 and then drops to virtually zero for type 5 households. One can also notice that the overall quality of the child care arrangement increases for all households benefiting from the possibility to deduct part of their child care expenses. Intuitively, the fact that the deduction is only granted to one-earner households is explained by the fact that, under the pure income tax optimum, for all skill types the overall quality of the child care arrangement is lower among one-earner household than among two-earner households. This implies that, in terms of the effects on social welfare, it is more valuable to raise the overall quality of child care arrangement for kids living in one-earner households. Moreover, given that under a pure income tax optimum the marginal tax rates are lower for two-earner households than for one-earner households, for the latter a deduction also represents a more effective instrument to boost the overall quality of the child care arrangement.

Comparing the rows "Optimal Tax" and "Deduction" of table 6, one can also see that supplementing an optimal nonlinear income tax with an optimal pair of deductibility rates (one for single-earner households and one for two-earner households) implies a further decrease in the proportion of two-earner households at the bottom of the skill distribution (for type 1 and type 2).

Finally, the welfare gains delivered by supplementing an optimal nonlinear income tax with an optimal pair of deductibility rates amount to 0.07% (calculated as the difference between the welfare gains reported in table 8 and the welfare gains reported in table 7: 4.01% – 3.94%).

Looking at the results displayed in table 9, the welfare gains increase to 0.25% when the deductibility rate is allowed to be income-dependent (calculated as the difference between the welfare gains reported in table 9 and the welfare gains reported in table 7: 4.19% – 3.94%). In this case the deductibility rate is 100% for all households of type 1 (irrespective of the mother's employment status) and then decreases monotonically to become 0 for households of type 4. Together with an increase in the marginal tax rate faced by type 1 households, this implies that the effective subsidy produced by the tax deductions reaches its maximum at the bottom of the skill distribution (being equal to 30.4% for type 1 single-earner households and 11.6% for type 1 two-earner households). Compared with the results obtained for the case of uniform deductibility rates, the optimal income-dependent deductibility rates raise the overall quality of the child care arrangement in households of type 1, 2 and 3.

Finally, comparing the rows "Deduction" and "Deduction (means-tested)" of table 6 , one can see that in terms of labor force participation, the only difference between the case when the deductibility rate is uniform and the case when it is income-dependent is that in the latter case the proportion of two-earner households slightly increases at the bottom of the skill distribution (but is still lower than under a pure income tax optimum). This happens despite the fact that the marginal income tax rate is negative (-1.7%) for two-earner households of type 1 when the deductibility rate is uniform and it becomes positive (11.6%) when the deductibility rate is allowed to be income-dependent. The increase in labor force participation is explained by the fact that in the former case there is no deduction that is granted to two-earner households whereas in the latter case two-earner households of type 1 are allowed to fully deduct their child care expenses.

6.3 Tax Credits

Let's now turn to the case when the government optimizes a set of tax credits alongside the optimal nonlinear income tax. The results for the case when the percentage of child care expenses that can be credited against the gross income tax only depends on the mother's employment status (and not on income) are displayed in table 10. The first thing to notice is that, as it happened for the case of uniform deductibility rates, the optimal subsidy is larger for one-earner households. In terms of optimal marginal income tax rates, the main difference with the case of uniform deductibility rates is that there is a significant drop in the marginal income tax rate faced by one-earner households of type 1. This suggests that when subsidies were granted through deductions the government was raising the marginal tax rate on type 1 households in order to raise the effective subsidy that they were receiving. In terms of effects on the overall quality of child care arrangement, one can also see that q is higher for all households when the government optimally chooses the set of uniform tax credit rates than when it optimally chooses the set of uniform deductibility rates. Instead, there is no difference between the two cases when one considers the percentage of one-earner households for each of the five skill types. Looking at the welfare gains, tax credits appear a better policy instrument to supplement the optimal nonlinear income tax: compared with the 0.07% delivered by uniform deductibility rates, the welfare gains associated with uniform tax credit rates amount to 0.49% (calculated as the difference between the welfare gains reported in table 10 and the welfare gains reported in table 7: 4.43% – 3.94%).

When allowing for the possibility of conditioning the tax credit rate both on the mother's employment status and the household income, the welfare gains rise to 0.72% (calculated as the difference between the welfare gains reported in table 11 and the welfare gains reported in table 7: 4.66% – 3.94%). In this case the optimal pattern of tax credit rates is monotonically decreasing in income for two-earner households (ranging from 45.6% to 3.9% and roughly mimicking the profile of subsidy rates that characterizes the benchmark allocation under the current US tax system; see table 4) and is also almost always monotonically decreasing in income for one-earner households (ranging from 49.7% to 6.2%). One again, however, for each skill type the tax credit rate turns out being more generous for one-earner couples than for two-earner couples (in contrast to what happens under the current US tax system).

Finally, notice that as it happened for the case of deductions, also in the case of tax credits there is an increase in the proportion of two-earner households when moving from a system where, conditional on the mother's employment status, tax credit rates are uniform to a system where they are allowed to be also income-dependent. And again, this happens despite the fact that the marginal income tax rate faced by two-earner households of type 1 increases by ten percentage points, from −5% to 5%. The reason why the increase in labor force participation occurs is that while two-earner households of type 1 get a tax credit rate of 15.2% in the uniform case, the rate increases for them to 45.6% in the income-dependent case.

Table 10: Uniform Tax Credit

Allocation in households where the mother works

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	TC
45.694	43.143	0.218	0.45	0.231	0.161	0.408	1.309	25.473	0.122	-0.05	15.2%
57.311	50.889	0.287	0.421	0.211	0.17	0.42	1.663	32.984	0.127	0.053	15.2%
70.001	58.803	0.318	0.403	0.201	0.176	0.423	2.046	40.81	0.129	0.132	15.2%
90.64	70.669	0.362	0.395	0.188	0.178	0.434	2.597	52.546	0.129	0.158	15.2%
148.873	102.513	0.455	0.425	0.16	0.169	0.471	3.846	81.924	0.126	0.014	15.2%

Allocation in households where the mother does not work

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	TC
36.545	33.14	0	0.447	0.299	0.162	0.339	1.606	26.637	0.155	0.159	35.8%
39.756	34.514	0	0.424	0.301	0.169	0.33	1.97	32.276	0.17	0.3	35.8%
46.094	37.347	0	0.427	0.303	0.168	0.329	2.345	38.559	0.174	0.344	35.8%
53.385	40.112	0	0.425	0.305	0.17	0.325	2.846	46.76	0.181	0.407	35.8%
86.588	55.558	0	0.541	0.306	0.134	0.359	4.064	71.127	0.175	0.052	35.8%

Welfare Gain = 4.43% of GDP

Rows correspond to household types 1-5. Taxable income M and consumption c expressed in thousands of USD (2006 values).

Table 11: Means-tested Tax Credit

Allocation in households where the mother works

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	TC
43.917	40.396	0.201	0.437	0.235	0.164	0.401	1.802	33.065	0.171	0.051	45.6%
57.093	49.806	0.285	0.42	0.21	0.17	0.42	1.916	37.246	0.147	0.075	30.6%
69.914	58.591	0.318	0.402	0.201	0.176	0.423	2.092	41.572	0.132	0.136	17.8%
90.542	71.415	0.362	0.394	0.189	0.179	0.433	2.405	49.145	0.12	0.15	5.5%
149.06	103.831	0.455	0.425	0.16	0.169	0.47	3.525	75.869	0.116	0.001	3.9%

Allocation in households where the mother does not work

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	TC
36.358	32.255	0	0.445	0.297	0.162	0.341	1.912	30.885	0.186	0.185	49.7%
39.938	34.582	0	0.426	0.301	0.169	0.331	1.983	32.496	0.171	0.294	36.4%
46.642	38.442	0	0.432	0.305	0.167	0.328	2.057	34.7	0.15	0.316	23.0%
54.448	42.672	0	0.433	0.311	0.168	0.321	2.161	37.527	0.132	0.355	6.2%
86.905	58.38	0	0.543	0.313	0.134	0.353	3.177	57.761	0.134	0.002	8.7%

Welfare Gain = 4.66% of GDP

Rows correspond to household types 1-5. Taxable income M and consumption c expressed in thousands of USD (2006 values).

6.4 Opting-out Public Provision Schemes

Tables 12-13 report the results characterizing an optimum when a nonlinear income tax is supplemented with an opting-out public provision scheme.[50] The welfare gains delivered by public provision amount to 0.88% for both variants of the opting-out provision scheme that we have considered (the value is obtained by subtracting the 3.94% welfare gain associated with the optimal nonlinear income tax reported in table 7 from the 4.82% welfare gain reported for the public provision schemes in tables 12-13),[51] implying that they are 0.16% larger than those obtained with income-dependent tax credit rates. Under both public provision schemes the quality chosen by the government for the publicly provided care is set at a level that is in between the one selected, at a pure income tax optimum, by households of type 3 and households of type 4 (compare tables 12-13 and table 7). Also the opting-in/opting-out pattern is the same under the two alternative schemes, with the first four households opting-in both among one-earner households and two-earner households, and the labor force participation is, for all skill types, identical to the one obtained with optimal income-dependent tax credit rates.

As one can see from tables 12-13, and in contrast to the results that we obtained when the weight on the externality term was set to zero (see tables 14-15), the generosity of each provision scheme does not differ markedly between one- and two-earner households. For the case of the first public provision scheme, households who opt-in get a subsidy on child care expenses that is either equal to 50.1%, if only one spouse works, or to 51.5% if both spouses work. For the case of the "impure" opting-out scheme, households who opt-in get either 0.346 hours of free child care, if only one spouse works, or 0.434 hours if both spouses work.[52] Of those households who opt-in, only two-earner households of type 4 use child care services for longer than the hours that are provided for free.

Finally, notice that of all the various subsidy schemes that we have considered, the opting-out public provision schemes are the most effective in reducing the dispersion in the overall quality of the child care arrangement for different households. Comparing tables 12-13 and table 7, one can see that this is achieved, without affecting the overall quality of child care arrangement in top-skilled households, by significantly boosting the quality in the lower part of the skill distribution.[53]

[50]From a computational point of view, these problems are more challenging to solve than the earlier problems as we will now also optimize over individuals' discrete binary decision to opt in or opt out of the publicly provided child care services.

[51]The welfare gain associated with the "impure" opting-out scheme is actually slightly larger, but the welfare gains from the two provision schemes are identical up to the second decimal.

[52]In this case the generosity of the system towards one- and two-earner households is roughly the same when one takes into account that two-earner households need more child care hours than one-earner households.

[53]This effect is slightly more pronounced in the case of the "impure" opting-out scheme.

Table 12: Public Provision I, $\overline{v} \neq 0$

Allocation in households where the mother works

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Opts in
43.65	40.26	0.198	0.435	0.236	0.165	0.399	2.017	36.399	0.192	6.0%	Yes
57.142	49.975	0.287	0.42	0.205	0.168	0.427	2.017	39.431	0.157	8.5%	Yes
70.369	59.437	0.322	0.404	0.193	0.17	0.437	2.017	41.192	0.13	14.2%	Yes
91.859	74.032	0.368	0.399	0.178	0.17	0.452	2.017	43.604	0.103	14.2%	Yes
149.906	103.179	0.458	0.428	0.16	0.169	0.472	3.399	73.636	0.111	-0.4%	No

Allocation in households where the mother does not work

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Opts in
36.082	32.345	0	0.441	0.299	0.164	0.337	2.017	32.084	0.196	19.1%	Yes
39.61	34.515	0	0.422	0.29	0.167	0.343	2.017	33.766	0.182	31.1%	Yes
46.037	38.068	0	0.426	0.282	0.165	0.353	2.017	35.783	0.161	34.8%	Yes
52.824	41.503	0	0.42	0.276	0.165	0.359	2.017	38.005	0.143	41.6%	Yes
87.359	58.391	0	0.546	0.315	0.134	0.351	2.946	54.354	0.123	-1.2%	No

Welfare Gain = 4.82% of GDP.
Subsidy rates: 51.5% (workers) and 50.1% (non-workers), PP quality level = 2.017

Rows correspond to household types 1-5. Taxable income M and consumption c expressed in thousands of USD (2006 values).

Table 13: Public Provision II, $\overline{v} \neq 0$

Allocation in households where the mother works

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Opts in
44.847	41.283	0.218	0.44	0.21	0.156	0.434	2.004	38.599	0.202	6%	Yes
57.932	50.807	0.293	0.424	0.201	0.165	0.434	2.004	39.641	0.156	6.5%	Yes
70.665	59.908	0.323	0.405	0.195	0.171	0.434	2.004	40.751	0.128	12.5%	Yes
91.631	74.098	0.366	0.399	0.185	0.176	0.439	2.004	42.597	0.1	12%	Yes
150.087	102.932	0.458	0.428	0.159	0.169	0.472	3.393	73.542	0.111	-0.4%	No

Allocation in households where the mother does not work

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Opts in
36.391	32.568	0	0.445	0.293	0.161	0.346	2.004	32.509	0.198	18%	Yes
39.645	34.627	0	0.423	0.287	0.167	0.346	2.004	33.799	0.182	31%	Yes
45.58	37.969	0	0.422	0.287	0.167	0.346	2.004	35.174	0.158	35.6%	Yes
51.932	41.182	0	0.413	0.285	0.169	0.346	2.004	37.031	0.139	42.9%	Yes
87.452	58.221	0	0.547	0.315	0.134	0.351	2.939	54.269	0.123	-1.2%	No

Welfare Gain = 4.82% of GDP.
Free hours: 0.434 (workers) and 0.346 (non-workers), PP quality level = 2.004

Rows correspond to household types 1-5. Taxable income M and consumption c expressed in thousands of USD (2006 values).

Table 14: Public Provision I, $\overline{v} = 0$

Allocation in households where the mother works

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Opts in
42.625	41.838	0.187	0.428	0.24	0.168	0.392	1.287	24.474	0.123	5.2%	Yes
56.501	51.982	0.282	0.416	0.208	0.169	0.423	1.287	26.807	0.1	6.5%	Yes
70.205	59.411	0.32	0.404	0.201	0.176	0.423	1.808	36.756	0.113	12%	No
90.25	70.913	0.36	0.393	0.189	0.179	0.432	2.295	47.154	0.114	15.8%	No
149.409	103.362	0.456	0.426	0.16	0.169	0.471	3.408	73.706	0.112	0%	No

Allocation in households where the mother does not work

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Opts in
35.7	34.808	0	0.436	0.285	0.163	0.352	1.287	22.835	0.132	16.7%	Yes
39.941	37.595	0	0.426	0.279	0.164	0.356	1.287	24.267	0.119	26.0%	Yes
46.61	41.644	0	0.432	0.275	0.162	0.362	1.287	25.827	0.104	29.1%	Yes
55.274	46.656	0	0.44	0.273	0.159	0.368	1.287	27.779	0.089	31.5%	Yes
86.918	58.394	0	0.543	0.315	0.135	0.35	2.951	54.355	0.124	0%	No

Welfare Gain = 4.63% of GDP.
Subsidy rates: 14.6% (workers) and 40.4% (non-workers), PP quality level = 1.287

Rows correspond to household types 1-5. Taxable income M and consumption c expressed in thousands of USD (2006 values).

Table 15: Public Provision II, $\overline{v} = 0$

Allocation in households where the mother works

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Opts in
42.667	42.024	0.186	0.429	0.244	0.168	0.387	1.26	23.862	0.119	4%	Yes
56.411	52.112	0.281	0.416	0.21	0.17	0.419	1.26	26.218	0.097	5.8%	Yes
70.134	59.388	0.319	0.403	0.202	0.176	0.423	1.808	36.748	0.113	12.1%	No
90.142	70.846	0.36	0.393	0.189	0.179	0.431	2.294	47.13	0.114	15.9%	No
149.422	103.339	0.456	0.426	0.16	0.169	0.471	3.407	73.697	0.112	0%	No

Allocation in households where the mother does not work

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$	Opts in
35.909	35.074	0	0.439	0.287	0.162	0.35	1.26	22.386	0.128	15.2%	Yes
40.019	37.834	0	0.427	0.284	0.165	0.35	1.26	23.664	0.115	25.1%	Yes
46.538	41.865	0	0.431	0.285	0.164	0.35	1.26	25.025	0.099	28.3%	Yes
55.21	46.964	0	0.439	0.287	0.162	0.35	1.26	26.865	0.083	30.3%	Yes
86.929	58.381	0	0.543	0.315	0.135	0.35	2.951	54.349	0.124	0%	No

Welfare Gain = 4.66% of GDP.
Free hours: 0.06 (workers) and 0.35 (non-workers), PP quality level = 1.26

Rows correspond to household types 1-5. Taxable income M and consumption c expressed in thousands of USD (2006 values).

7 Summary and conclusion

In this paper we have studied, in the context of a Mirrleesian optimal tax model, the relative merits of some of the most common types of child care subsidies: tax deductions, tax credits, and opting-out public provision schemes. The point of departure for the analysis has been that there is a special tax treatment for families with children in child care ages. Thus, in our quantitative investigation of child care subsidies we have kept the net tax revenue from this group of taxpayers constant.

We have designed our simulation model in such a way that it incorporates important aspects of the US economy. We have considered a joint system of taxation and employed wage distributions calibrated to fit the empirical wage distributions of mothers and fathers with kids in child care age, and we have disciplined our parameters using time-use data obtained from the American Time Use Survey.

While previous studies in the optimal tax literature have analyzed the welfare effects of child care subsidies in models where the quality of center-based child care was treated as exogenously given, so that child care expenditures were only a function of the number of hours spent by a kid in a child care center, we have allowed parents to choose both the quantity and the quality of center-based child care services. Moreover, we have allowed both the time spent by kids at a child care facility and the time spent by parents with their offspring to affect the overall quality of the child care arrangement, and therefore contribute to the human capital development of children.

In our model there are two potential mechanisms by which subsidies to child care can be welfare-enhancing. One is to mitigate binding self-selection constraints, thereby allowing to achieve the desired redistributive goals at lower efficiency costs. The other is to increase the overall quality of child care arrangements, thereby fostering the human capital development in children. In particular, we have used an externality term in the social welfare function to capture the idea that in a good society all children should have equal chances in life (no one should be left behind) and that one way to move in this direction is to let all children have access to good quality child care. This is a view held by many proponents of subsidized child care. For countries who engage heavily in income redistribution, it might be more important to mitigate the distortionary effects associated with income taxation. In other countries that engage less in income redistribution, the externality argument in favor of subsidized child care might carry more weight.

In contrast to what has been obtained in previous optimal taxation studies, our results indicate that there is no scope for using child care subsidies to reduce the inefficiencies associated with income redistribution, at least not when the subsidies are implemented via tax deductions or refundable tax credits, i.e. when the subsidies are based on the child care expenditures incurred by households. The main reason for this result is that, in contrast to previous models, we allow parents to choose both the quality and the quantity of child care services that they wish to purchase.

In a model where the quality of child care is fixed, the variation in child care expenditure is largely driven by variation in child care hours, which is strongly correlated with hours of work. This implies that if a high-skilled household were to mimic a low-skilled one, the expenditure on child care services would be higher for the low-skilled than for the high-skilled mimicker. In our model, instead, a high expenditure can be the result of either a high quality of the chosen child care facility or a high number of child care hours. If a high-skilled mimicker chooses a higher quality of child care services than a low-skilled, it is then conceivable that child care expenditures are larger for the former. This undermines the role for child care subsidies as a mimicking-deterring device. However, subsidies delivered through an opting-out public provision scheme remain a useful instrument for mimicking-deterring purposes. The reason is that under an opting-out public provision scheme the quality of the center-based care is set by the government and is no longer a choice variable for the households who decide to opt-in. Granting subsidies only to households who opt in implies then that the private value of a subsidy is only a function of the number of child care hours that are used. This in turn implies that, when a low-skilled household opts-in, a subsidy to child care expenditures yields a larger benefit to a low-skilled household than to a mimicker unless the latter also opts-in and at the same time demands more hours of center-based care than a true low-skilled (which is unlikely to happen given that a high-skilled mimicker needs to work fewer hours than a low-skilled).

Even though child care subsidies, either in the form of tax deductions or refundable tax credits, do not necessarily seem to reduce the distortions associated with income taxation, they can be welfare-enhancing devices when viewed as externality-correcting instruments. In particular, focusing on the case when the government maximizes a utilitarian social welfare function, they deliver welfare gains ranging from 0.07% to 0.72%. However, even larger welfare gains (0.88%) are obtained by using an opting-out public provision scheme. Of course, these numbers are only meant to be illustrative, as we have made no attempts to empirically quantify the appropriate weight that externalities associated with human capital investment in children should carry in the social objective.

Restricting attention to the relative merits of tax deductions and refundable tax credits, our results indicate that the latter deliver larger welfare gains. The lowest welfare gains that we obtain (0.07%) are obtained for the case when households can deduct child care expenses against their earned income and the deductibility rate is only differentiated according to the mother's employment status (one-earner households versus two-earner households). The largest welfare gains (0.72%) are instead obtained for the case when households get refundable tax credits for their child care expenses and the tax credit rates are differentiated both according to the mother's employment status and the household income. In the latter case the optimal profile of child care subsidies for two-earner households roughly mimics the one prevailing under the current US tax system. However, in contrast to the current practice characterizing the US tax system, more generous subsidies should be granted to one-earner households than to two-earner households.

Finally, we would like to mention a potentially broader implication of our results. One of the most frequently occurring results in applied tax policy discussion is the Corlett and Hague (1953) recommendation that goods complementary to leisure should be taxed (or, equivalently, that goods complementary to labor should be subsidized) to reduce the inefficiencies associated with income taxation. We have highlighted that in an optimal income tax framework, with explicit redistributional objectives, such results need to be qualified to take into account the quality dimension of the goods in question. Specifically, the desirability of taxes and subsidies depends on the pattern of expenditure between mimickers and true-types in the optimal income tax problem. To be able to derive meaningful policy insights, these expenditure patterns must be assessed through the lens of a calibrated model.

A Proofs and derivations

A.1 Derivation of results stated in Proposition 1

From the Lagrangian of the government's problem:

$$\begin{aligned}
\mathcal{L} &= \theta^1 V^1\left(M^1, B^1, \alpha, \beta\right) + \theta^2 V^2\left(M^2, B^2, \alpha, \beta\right) + \nu\left(q^1, q^2\right) \\
&\quad + \lambda\left[V^2\left(M^2, B^2, \alpha, \beta\right) - V^2\left(M^1, B^1, \alpha, \beta\right)\right] \\
&\quad + \mu\left\{\left(M^1 - B^1\right)\pi^1 + \left(M^2 - B^2\right)\pi^2 - \left[\pi^1 p\left(q_c^1\right) h_c^1 + \pi^2 p\left(q_c^2\right) h_c^2\right]\beta - R\right\} \\
&\quad + \delta\left[1 - \alpha - \beta\right],
\end{aligned}$$

we can derive the following set of first order conditions:

$$\frac{\partial \mathcal{L}}{\partial M^2} = \left(\theta^2 + \lambda\right)\frac{\partial V^2}{\partial M^2} + \frac{\partial \nu\left(q^1, q^2\right)}{\partial q^2}\frac{dq^2}{dM^2} + \mu\pi^2\left(1 - \beta\frac{dD^2}{dM^2}\right) = 0 \tag{B1}$$

$$\frac{\partial \mathcal{L}}{\partial B^2} = \left(\theta^2 + \lambda\right)\frac{\partial V^2}{\partial B^2} + \frac{\partial \nu\left(q^1, q^2\right)}{\partial q^2}\frac{dq^2}{dB^2} - \mu\pi^2\left(1 + \beta\frac{dD^2}{dB^2}\right) = 0 \tag{B2}$$

$$\frac{\partial \mathcal{L}}{\partial M^1} = \theta^1\frac{\partial V^1}{\partial M^1} - \lambda\frac{\partial \widehat{V}^2}{\partial M^1} + \frac{\partial \nu\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{dM^1} + \mu\pi^1\left(1 - \beta\frac{dD^1}{dM^1}\right) = 0 \tag{B3}$$

$$\frac{\partial \mathcal{L}}{\partial B^1} = \theta^1\frac{\partial V^1}{\partial B^1} - \lambda\frac{\partial \widehat{V}^2}{\partial B^1} + \frac{\partial \nu\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{dB^1} - \mu\pi^1\left(1 + \beta\frac{dD^1}{dB^1}\right) = 0 \tag{B4}$$

and

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \theta^1 \frac{\partial V^1}{\partial \alpha} + \left(\theta^2 + \lambda\right) \frac{\partial V^2}{\partial \alpha} - \lambda \frac{\partial \widehat{V}^2}{\partial \alpha} + \sum_{i=1}^{2} \frac{\partial v\left(q^1, q^2\right)}{\partial q^i} \frac{dq^i}{d\alpha} - \mu\beta \sum_{i=1}^{2} \pi^i \frac{dD^i}{d\alpha} - \delta = 0 \tag{B5}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \beta} &= \theta^1 \frac{\partial V^1}{\partial \beta} + \left(\theta^2 + \lambda\right) \frac{\partial V^2}{\partial \beta} - \lambda \frac{\partial \widehat{V}^2}{\partial \beta} + \sum_{i=1}^{2} \frac{\partial v\left(q^1, q^2\right)}{\partial q^i} \frac{dq^i}{d\beta} \\ &\quad -\mu \left[\sum_{i=1}^{2} \pi^i \left(D^i + \beta \frac{dD^i}{d\beta} \right) \right] - \delta = 0, \end{aligned} \tag{B6}$$

where we have used a " $\widehat{\ }$ " symbol to denote a variable when pertaining to a mimicker.

Using (12) to substitute in the first order condition for (B6) gives:

$$\begin{aligned} &\theta^1 D^1 \frac{\partial u^1}{\partial c} + \left(\theta^2 + \lambda\right) D^2 \frac{\partial u^2}{\partial c} - \lambda \widehat{D}^2 \frac{\partial \widehat{u}^2}{\partial c} + \sum_{i=1}^{2} \frac{\partial v\left(q^1, q^2\right)}{\partial q^i} \frac{dq^i}{d\beta} - \\ &\qquad - \mu \left[\sum_{i=1}^{2} \pi^i \left(D^i + \beta \frac{dD^i}{d\beta} \right) \right] - \delta = 0 \end{aligned} \tag{B7}$$

Using the Slutsky equation to decompose $dq^i/d\beta$ and $dD^i/d\beta$ into a substitution and an income effect allows rewriting (B7) as:

$$\begin{aligned} &\theta^1 D^1 \frac{\partial u^1}{\partial c} + \left(\theta^2 + \lambda\right) D^2 \frac{\partial u^2}{\partial c} - \lambda \widehat{D}^2 \frac{\partial \widehat{u}^2}{\partial c} + \sum_{i=1}^{2} \frac{\partial v\left(q^1, q^2\right)}{\partial q^i} \left(\frac{d\widetilde{q}^i}{d\beta} + D^i \frac{dq^i}{dB^i} \right) \\ &-\mu \left[\sum_{i=1}^{2} \pi^i \left(D^i + \beta \frac{d\widetilde{D}^i}{d\beta} + \beta D^i \frac{dD^i}{dB^i} \right) \right] - \delta \\ =\ & 0 \end{aligned} \tag{B8}$$

Multiplying all terms in (B2) by $-D^2$ and all terms in (B4) by $-D^1$ and then adding the resulting expressions to (B8) gives:

$$\begin{aligned} &-\left(\theta^2 + \lambda\right) \frac{\partial V^2}{\partial B^2} D^2 - \frac{\partial v\left(q^1, q^2\right)}{\partial q^2} \frac{dq^2}{dB^2} D^2 + \mu \pi^2 \left(1 + \beta \frac{dD^2}{dB^2} \right) D^2 \\ &-\theta^1 \frac{\partial V^1}{\partial B^1} D^1 + \lambda \frac{\partial \widehat{V}^2}{\partial B^1} D^1 - \frac{\partial v\left(q^1, q^2\right)}{\partial q^1} \frac{dq^1}{dB^1} D^1 + \mu \pi^1 \left(1 + \beta \frac{dD^1}{dB^1} \right) D^1 \\ &+\theta^1 D^1 \frac{\partial u^1}{\partial c} + \left(\theta^2 + \lambda\right) D^2 \frac{\partial u^2}{\partial c} - \lambda \widehat{D}^2 \frac{\partial \widehat{u}^2}{\partial c} + \sum_{i=1}^{2} \frac{\partial v\left(q^1, q^2\right)}{\partial q^i} \left(\frac{d\widetilde{q}^i}{d\beta} + D^i \frac{dq^i}{dB^i} \right) \\ &-\mu \left[\sum_{i=1}^{2} \pi^i \left(D^i + \beta \frac{d\widetilde{D}^i}{d\beta} + \beta D^i \frac{dD^i}{dB^i} \right) \right] - \delta \\ =\ & 0 \end{aligned}$$

Simplifying terms in the equation above gives:

$$\lambda \frac{\partial \widehat{V}^2}{\partial B^1}\left(D^1 - \widehat{D}^2\right) + \sum_{i=1}^{2} \frac{\partial v\left(q^1, q^2\right)}{\partial q^i}\frac{d\widetilde{q}^i}{d\beta} - \mu\beta\left[\sum_{i=1}^{2}\pi^i \frac{d\widetilde{D}^i}{d\beta}\right] - \delta = 0 \tag{B9}$$

Rearranging (B9) gives (15).

Using (11) to substitute in the first order condition for (B5) gives:

$$\begin{aligned} & \theta^1 D^1 T'\left(M^1\right)\frac{\partial u^1}{\partial c} + \left(\theta^2 + \lambda\right) D^2 T'\left(M^2\right)\frac{\partial u^2}{\partial c} - \lambda \widehat{D}^2 \widehat{T}'\left(M^1\right)\frac{\partial \widehat{u}^2}{\partial c} \\ & + \sum_{i=1}^{2}\frac{\partial v\left(q^1, q^2\right)}{\partial q^i}\frac{dq^i}{d\alpha} - \mu\beta\sum_{i=1}^{2}\pi^i \frac{dD^i}{d\alpha} - \delta \\ = \ & 0 \end{aligned} \tag{B10}$$

Using the Slutsky equation to decompose $dq^i/d\alpha$ and $dD^i/d\alpha$ into a substitution and an income effect allows rewriting (B10) as:

$$\begin{aligned} & \theta^1 D^1 T'\left(M^1\right)\frac{\partial u^1}{\partial c} + \left(\theta^2 + \lambda\right) D^2 T'\left(M^2\right)\frac{\partial u^2}{\partial c} - \lambda \widehat{D}^2 \widehat{T}'\left(M^1\right)\frac{\partial \widehat{u}^2}{\partial c} \\ & + \sum_{i=1}^{2}\frac{\partial v\left(q^1, q^2\right)}{\partial q^i}\left(\frac{d\widetilde{q}^i}{d\alpha} + D^i T'\left(M^i\right)\frac{dq^i}{dB^i}\right) - \mu\beta\sum_{i=1}^{2}\pi^i\left(\frac{d\widetilde{D}^i}{d\alpha} + D^i T'\left(M^i\right)\frac{dD^i}{dB^i}\right) - \delta \\ = \ & 0 \end{aligned} \tag{B11}$$

Multiplying all terms in (B2) by $-T'\left(M^2\right)D^2$ and all terms in (B4) by $-T'\left(M^1\right)D^1$ and then adding the resulting expressions to (B11) gives:

$$\begin{aligned} & -\left(\theta^2 + \lambda\right)\frac{\partial V^2}{\partial B^2}T'\left(M^2\right)D^2 - \frac{\partial v\left(q^1, q^2\right)}{\partial q^2}\frac{dq^2}{dB^2}T'\left(M^2\right)D^2 + \mu\pi^2\left(1 + \beta\frac{dD^2}{dB^2}\right)T'\left(M^2\right)D^2 \\ & -\theta^1\frac{\partial V^1}{\partial B^1}T'\left(M^1\right)D^1 + \lambda\frac{\partial \widehat{V}^2}{\partial B^1}T'\left(M^1\right)D^1 - \frac{\partial v\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{dB^1}T'\left(M^1\right)D^1 \\ & +\mu\pi^1\left(1 + \beta\frac{dD^1}{dB^1}\right)T'\left(M^1\right)D^1 \\ & +\theta^1 D^1 T'\left(M^1\right)\frac{\partial u^1}{\partial c} + \left(\theta^2 + \lambda\right) D^2 T'\left(M^2\right)\frac{\partial u^2}{\partial c} - \lambda \widehat{D}^2 \widehat{T}'\left(M^1\right)\frac{\partial \widehat{u}^2}{\partial c} \\ & + \sum_{i=1}^{2}\frac{\partial v\left(q^1, q^2\right)}{\partial q^i}\left(\frac{d\widetilde{q}^i}{d\alpha} + D^i T'\left(M^i\right)\frac{dq^i}{dB^i}\right) - \mu\beta\sum_{i=1}^{2}\pi^i\left(\frac{d\widetilde{D}^i}{d\alpha} + D^i T'\left(M^i\right)\frac{dD^i}{dB}\right) - \delta = 0 \end{aligned}$$

Simplifying terms in the equation above gives:

$$\lambda\frac{\partial\widehat{V}^2}{\partial B^1}\left[T'\left(M^1\right)D^1-\widehat{D}^2\widehat{T}'\left(M^1\right)\right]-\mu\beta\sum_{i=1}^{2}\pi^i\frac{d\widetilde{D}^i}{d\alpha}+\mu\sum_{i=1}^{2}\pi^iT'\left(M^i\right)D^i-$$
$$-\delta+\sum_{i=1}^{2}\frac{\partial v\left(q^1,q^2\right)}{\partial q^i}\frac{d\widetilde{q}^i}{d\alpha}=0 \tag{B12}$$

Rearranging (B12) gives (14).

To derive the optimal marginal income tax rate faced by a member of a type 2 household, rewrite (B1) and (B2) as:

$$\left(\theta^2+\lambda\right)\frac{\partial V^2}{\partial M^2} = -\frac{\partial v\left(q^1,q^2\right)}{\partial q^2}\frac{dq^2}{dM^2}-\mu\pi^2\left(1-\beta\frac{dD^2}{dM^2}\right) \tag{B13}$$
$$\left(\theta^2+\lambda\right)\frac{\partial V^2}{\partial B^2} = -\frac{\partial v\left(q^1,q^2\right)}{\partial q^2}\frac{dq^2}{dB^2}+\mu\pi^2\left(1+\beta\frac{dD^2}{dB^2}\right) \tag{B14}$$

Dividing (B13) by (B14) and multiplying the result by the right hand side of (B14) gives:

$$\frac{\frac{\partial V^2}{\partial M^2}}{\frac{\partial V^2}{\partial B^2}}\left[-\frac{\partial v\left(q^1,q^2\right)}{\partial q^2}\frac{dq^2}{dB^2}+\mu\pi^2\left(1+\beta\frac{dD^2}{dB^2}\right)\right]=-\frac{\partial v\left(q^1,q^2\right)}{\partial q^2}\frac{dq^2}{dM^2}-\mu\pi^2\left(1-\beta\frac{dD^2}{dM^2}\right),$$

which can be rewritten as:

$$\left(1+\frac{\frac{\partial V^2}{\partial M^2}}{\frac{\partial V^2}{\partial B^2}}\right)\mu\pi^2=\frac{\frac{\partial V^2}{\partial M^2}}{\frac{\partial V^2}{\partial B^2}}\frac{\partial v\left(q^1,q^2\right)}{\partial q^2}\frac{dq^2}{dB^2}-\frac{\partial v\left(q^1,q^2\right)}{\partial q^2}\frac{dq^2}{dM^2}+\mu\pi^2\beta\frac{dD^2}{dM^2}-\frac{\frac{\partial V^2}{\partial M^2}}{\frac{\partial V^2}{\partial B^2}}\mu\pi^2\beta\frac{dD^2}{dB^2}$$

Using (8) allows expressing the optimal marginal income tax rate faced by a member of a type 2 household as:

$$T'\left(M^2\right)=\beta\left(\frac{dD^2}{dM^2}\right)_{dV^2=0}-\frac{\frac{\partial v\left(q^1,q^2\right)}{\partial q^2}}{\mu\pi^2}\left(\frac{dq^2}{dM^2}\right)_{dV^2=0} \tag{B15}$$

A similar procedure applied on (B3) and (B4) allows deriving the following expression for the optimal marginal income tax rate faced by a member of a type 1 household:

$$T'\left(M^1\right)=\frac{\lambda\frac{\partial\widehat{V}^2}{\partial B^1}}{\mu\pi^1}\left[MRS^1_{MB}-\widehat{MRS}_{MB}\right]+\beta\left(\frac{dD^1}{dM^1}\right)_{dV^1=0}-\frac{\frac{\partial v\left(q^1,q^2\right)}{\partial q^1}}{\mu\pi^1}\left(\frac{dq^1}{dM^1}\right)_{dV^1=0} \tag{B16}$$

For $\beta = 0$, (B16) and (B15) simplify to:

$$
\begin{aligned}
T'\left(M^2\right) &= -\frac{\frac{\partial v\left(q^1,q^2\right)}{\partial q^2}}{\mu\pi^2}\left(\frac{dq^2}{dM^2}\right)_{dV^2=0} && \text{(B17)}\\
T'\left(M^1\right) &= \frac{\lambda\frac{\partial\widehat{V}^2}{\partial B^1}}{\mu\pi^1}\left[MRS^{\,1}_{MB} - \widehat{MRS}_{MB}\right] - \frac{\frac{\partial v\left(q^1,q^2\right)}{\partial q^1}}{\mu\pi^1}\left(\frac{dq^1}{dM^1}\right)_{dV^1=0} && \text{(B18)}
\end{aligned}
$$

Since (B18) can be rewritten in an equivalent way as:

$$
\begin{aligned}
T'\left(M^1\right) &= \frac{\lambda\frac{\partial\widehat{V}^2}{\partial B^1}}{\mu\pi^1}\left[1 - T'\left(M^1\right) - 1 + \widehat{T}'\left(M^1\right)\right] - \frac{\frac{\partial v\left(q^1,q^2\right)}{\partial q^1}}{\mu\pi^1}\left(\frac{dq^1}{dM^1}\right)_{dV^1=0}\\
&= \frac{\lambda\frac{\partial\widehat{V}^2}{\partial B^1}}{\mu\pi^1}\left[\widehat{T}'\left(M^1\right) - T'\left(M^1\right)\right] - \frac{\frac{\partial v\left(q^1,q^2\right)}{\partial q^1}}{\mu\pi^1}\left(\frac{dq^1}{dM^1}\right)_{dV^1=0}, && \text{(B19)}
\end{aligned}
$$

when $\beta = 0$ one can substitute (B19) and (B17) into (B12) to obtain:

$$
\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\alpha} &= \lambda\frac{\partial\widehat{V}^2}{\partial B^1}\left[T'\left(M^1\right)D^1 - \widehat{D}^2\widehat{T}'\left(M^1\right)\right] - \delta + \sum_{i=1}^{2}\frac{\partial v\left(q^1,q^2\right)}{\partial q^i}\frac{d\widetilde{q}^i}{d\alpha}\\
&\quad + D^1\left\{\lambda\frac{\partial\widehat{V}^2}{\partial B^1}\left[\widehat{T}'\left(M^1\right) - T'\left(M^1\right)\right] - \frac{\partial v\left(q^1,q^2\right)}{\partial q^1}\left(\frac{dq^1}{dM^1}\right)_{dV^1=0}\right\}\\
&\quad - D^2\frac{\partial v\left(q^1,q^2\right)}{\partial q^2}\left(\frac{dq^2}{dM^2}\right)_{dV^2=0}
\end{aligned}
$$

Simplifying terms in the above equation gives (16).

A.2 Derivations of results stated in Proposition 2

From the Lagrangian of the government's problem:

$$
\begin{aligned}
\mathcal{L} &= \theta^1 V^{1in}\left(M^1, B^1, \overline{q}_c, s\right) + \theta^2 V^{2out}\left(M^2, B^2\right) + v\left(q^1, q^2\right)\\
&\quad + \gamma^1\left[V^{1in}\left(M^1, B^1, \overline{q}_c, s\right) - V^{1out}\left(M^1, B^1\right)\right]\\
&\quad + \gamma^2\left[V^{2out}\left(M^2, B^2\right) - V^{2in}\left(M^2, B^2, \overline{q}_c, s\right)\right]\\
&\quad + \lambda^{out}\left[V^{2out}\left(M^2, B^2\right) - V^{2out}\left(M^1, B^1\right)\right]\\
&\quad + \lambda^{in}\left[V^{2out}\left(M^2, B^2\right) - V^{2in}\left(M^1, B^1, \overline{q}_c, s\right)\right]\\
&\quad + \mu\left\{\left(M^1 - B^1\right)\pi^1 + \left(M^2 - B^2\right)\pi^2 - \pi^1 sp\left(\overline{q}_c\right)h_c^{1in} - R\right\},
\end{aligned}
$$

the first order conditions with respect to B^1, $\overline{q}_c$ and s are respectively given by:

$$\frac{\partial\mathcal{L}}{\partial B^1}=\left(\theta^1+\gamma^1\right)\frac{\partial V^{1in}}{\partial B^1}-\gamma^1\frac{\partial V^{1out}}{\partial B^1}-\lambda^{out}\frac{\partial\widehat{V}^{2out}}{\partial B^1}-\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}+\frac{\partial\nu\left(q^1,q^2\right)}{\partial q^1}\frac{dq^1}{dB^1}- \\ -\mu\pi^1\left[1+sp\left(\overline{q}_c\right)\frac{\partial h_c^{1in}}{\partial B^1}\right]=0, \quad \text{(B20)}$$

$$\frac{\partial\mathcal{L}}{\partial\overline{q}_c}=\left(\theta^1+\gamma^1\right)\frac{\partial V^{1in}}{\partial\overline{q}_c}-\gamma^2\frac{\partial V^{2in}}{\partial\overline{q}_c}-\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial\overline{q}_c}+\frac{\partial\nu\left(q^1,q^2\right)}{\partial q^1}\frac{dq^1}{d\overline{q}_c}- \\ -\mu\pi^1 s\left[h_c^{1in}p'\left(\overline{q}_c\right)+p\left(\overline{q}_c\right)\frac{\partial h_c^{1in}}{\partial\overline{q}_c}\right]=0, \quad \text{(B21)}$$

$$\frac{\partial\mathcal{L}}{\partial s}=\left(\theta^1+\gamma^1\right)\frac{\partial V^{1in}}{\partial s}-\gamma^2\frac{\partial V^{2in}}{\partial s}-\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial s}+\frac{\partial\nu\left(q^1,q^2\right)}{\partial q^1}\frac{dq^1}{ds}- \\ -\mu\pi^1 p\left(\overline{q}_c\right)\left[h_c^{1in}+s\frac{\partial h_c^{1in}}{\partial s}\right]=0. \quad \text{(B22)}$$

Adding and subtracting $\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\frac{\partial V^{1in}}{\partial\overline{q}_c}/\frac{\partial V^{1in}}{\partial B^1}$ to (B21), one can rewrite it as:

$$\left[\left(\theta^1+\gamma^1\right)\frac{\partial V^{1in}}{\partial B^1}-\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\right]\frac{\frac{\partial V^{1in}}{\partial\overline{q}_c}}{\frac{\partial V^{1in}}{\partial B^1}}+\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\left(\frac{\frac{\partial V^{1in}}{\partial\overline{q}_c}}{\frac{\partial V^{1in}}{\partial B^1}}-\frac{\frac{\partial\widehat{V}^{2in}}{\partial\overline{q}_c}}{\frac{\partial\widehat{V}^{2in}}{\partial B^1}}\right) \\ -\gamma^2\frac{\partial V^{2in}}{\partial\overline{q}_c}+\frac{\partial\nu\left(q^1,q^2\right)}{\partial q^1}\frac{dq^1}{d\overline{q}_c}-\mu\pi^1 s\left[h_c^{1in}p'\left(\overline{q}_c\right)+p\left(\overline{q}_c\right)\frac{\partial h_c^{1in}}{\partial\overline{q}_c}\right]=0.$$

From the first order condition (B20) one can derive an expression for $\left(\theta^1+\gamma^1\right)\frac{\partial V^{1in}}{\partial B^1}-\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}$ that can be substituted in (B23) to obtain:

$$\left\{\gamma^1\frac{\partial V^{1out}}{\partial B^1}+\lambda^{out}\frac{\partial\widehat{V}^{2out}}{\partial B^1}-\frac{\partial\nu\left(q^1,q^2\right)}{\partial q^1}\frac{dq^1}{dB^1}+\mu\pi^1\left[1+sp\left(\overline{q}_c\right)\frac{\partial h_c^{1in}}{\partial B^1}\right]\right\}\frac{\frac{\partial V^{1in}}{\partial\overline{q}_c}}{\frac{\partial V^{1in}}{\partial B^1}}+ \\ \lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\left(\frac{\frac{\partial V^{1in}}{\partial\overline{q}_c}}{\frac{\partial V^{1in}}{\partial B^1}}-\frac{\frac{\partial\widehat{V}^{2in}}{\partial\overline{q}_c}}{\frac{\partial\widehat{V}^{2in}}{\partial B^1}}\right)-\gamma^2\frac{\partial V^{2in}}{\partial\overline{q}_c}+\frac{\partial\nu\left(q^1,q^2\right)}{\partial q^1}\frac{dq^1}{d\overline{q}_c}- \\ -\mu\pi^1 s\left[h_c^{1in}p'\left(\overline{q}_c\right)+p\left(\overline{q}_c\right)\frac{\partial h_c^{1in}}{\partial\overline{q}_c}\right]=0. \quad \text{(B23)}$$

Denoting by $MRS^{1in}_{\overline{q}_cB}$ and $\widehat{MRS}^{2in}_{\overline{q}_cB}$ the marginal rate of substitution between $\overline{q}_c$ and B for, respectively, an opting-in type-1 household and an opting-in type-2 mimicker, one can rewrite

(B23) as:

$$\mu\pi^1\left\{MRS^{1in}_{\overline{q}_cB} - s\left[h_c^{1in}p'(\overline{q}_c) + p(\overline{q}_c)\left(\frac{\partial h_c^{1in}}{\partial\overline{q}_c}\right)_{dV^{1in}=0}\right]\right\}$$
$$+\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\left(MRS^{1in}_{\overline{q}_cB} - \widehat{MRS}^{2in}_{\overline{q}_cB}\right) + \lambda^{out}\frac{\partial\widehat{V}^{2out}}{\partial B^1}MRS^{1in}_{\overline{q}_cB} + \frac{\partial v\left(q^1,q^2\right)}{\partial q^1}\left(\frac{dq^1}{d\overline{q}_c}\right)_{dV^{1in}=0}$$
$$-\gamma^2\frac{\partial V^{2in}}{\partial\overline{q}_c} + \gamma^1\frac{\partial V^{1out}}{\partial B^1}MRS^{1in}_{\overline{q}_cB} = 0,$$

and thus, rearranging terms, one immediately obtains the result stated by (17).

Taking into account that $\frac{\partial V^{1in}}{\partial s} = p(\overline{q}_c)\,h_c^{1in}\frac{\partial V^{1in}}{\partial B^1}$, $\frac{\partial V^{2in}}{\partial s} = p(\overline{q}_c)\,h_c^{2in}\frac{\partial V^{2in}}{\partial B^2}$, $\frac{\partial\widehat{V}^{2in}}{\partial s} = p(\overline{q}_c)\,\widehat{h}_c^{2in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}$, the first order condition (B22) can be rewritten as:

$$\left(\theta^1+\gamma^1\right)p(\overline{q}_c)\,h_c^{1in}\frac{\partial V^{1in}}{\partial B^1} - \gamma^2 p(\overline{q}_c)\,h_c^{2in}\frac{\partial V^{2in}}{\partial B^2} - \lambda^{in}p(\overline{q}_c)\,\widehat{h}_c^{2in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}$$
$$+\frac{\partial v\left(q^1,q^2\right)}{\partial q^1}\frac{dq^1}{ds} - \mu\pi^1 p(\overline{q}_c)\left[h_c^{1in} + s\frac{\partial h_c^{1in}}{\partial s}\right] = 0.$$

Multiplying (B20) by $-p(\overline{q}_c)\,h_c^{1in}$ and adding the resulting expression to (B24) gives:

$$-\left(\theta^1+\gamma^1\right)\frac{\partial V^{1in}}{\partial B^1}p(\overline{q}_c)\,h_c^{1in} + \gamma^1\frac{\partial V^{1out}}{\partial B^1}p(\overline{q}_c)\,h_c^{1in} + \lambda^{out}\frac{\partial\widehat{V}^{2out}}{\partial B^1}p(\overline{q}_c)\,h_c^{1in}$$
$$+\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}p(\overline{q}_c)\,h_c^{1in} - \frac{\partial v\left(q^1,q^2\right)}{\partial q^1}\frac{dq^1}{dB^1}p(\overline{q}_c)\,h_c^{1in} + \mu\pi^1\left[1 + sp(\overline{q}_c)\frac{\partial h_c^{1in}}{\partial B^1}\right]p(\overline{q}_c)\,h_c^{1in}$$
$$+\left(\theta^1+\gamma^1\right)p(\overline{q}_c)\,h_c^{1in}\frac{\partial V^{1in}}{\partial B} - \gamma^2 p(\overline{q}_c)\,h_c^{2in}\frac{\partial V^{2in}}{\partial B^2} - \lambda^{in}p(\overline{q}_c)\,\widehat{h}_c^{2in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}$$
$$+\frac{\partial v\left(q^1,q^2\right)}{\partial q^1}\frac{dq^1}{ds} - \mu\pi^1 p(\overline{q}_c)\left[h_c^{1in} + s\frac{\partial h_c^{1in}}{\partial s}\right] = 0.$$

Using the Slutsky equation to decompose $\partial h_c^{1in}/\partial s$ into a substitution and an income effect, and simplifying and rearranging terms in the equation above gives:

$$\left[\gamma^1\frac{\partial V^{1out}}{\partial B^1}h_c^{1in} - \gamma^2 h_c^{2in}\frac{\partial V^{2in}}{\partial B^2}\right]p(\overline{q}_c) + \left[\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\left(h_c^{1in} - \widehat{h}_c^{2in}\right) + \lambda^{out}\frac{\partial\widehat{V}^{2out}}{\partial B^1}h_c^{1in}\right]p(\overline{q}_c)$$
$$+\frac{\partial v\left(q^1,q^2\right)}{\partial q^1}\left(\frac{dq^1}{ds}\right)_{dV^{1in}=0} = \mu\pi^1 p(\overline{q}_c)\,s\frac{\partial\widetilde{h}_c^{1in}}{\partial s}, \quad \text{(B24)}$$

where $\widetilde{h}_c^{1in}$ denotes the compensated (Hicksian) demand for hours of center-based child care by low-skilled households.

From (B24) one can immediately obtain the result stated in (19).

A.3 Derivations of results stated in Proposition 3

From the Lagrangian of the government's problem:

$$\begin{aligned}
\mathcal{L} \;=\; & \theta^1 V^{1in}\left(M^1, B^1, \overline{q}_c, \overline{h}_c\right) + \theta^2 V^{2out}\left(M^2, B^2\right) + \nu\left(q^1, q^2\right) \\
& +\gamma^1\left[V^{1in}\left(M^1, B^1, \overline{q}_c, \overline{h}_c\right) - V^{1out}\left(M^1, B^1\right)\right] \\
& +\gamma^2\left[V^{2out}\left(M^2, B^2\right) - V^{2in}\left(M^2, B^2, \overline{q}_c, \overline{h}_c\right)\right] \\
& +\lambda^{out}\left[V^{2out}\left(M^2, B^2\right) - V^{2out}\left(M^1, B^1\right)\right] \\
& +\lambda^{in}\left[V^{2out}\left(M^2, B^2\right) - V^{2in}\left(M^1, B^1, \overline{q}_c, \overline{h}_c\right)\right] \\
& +\mu\left\{\left(M^1 - B^1\right)\pi^1 + \left(M^2 - B^2\right)\pi^2 - \pi^1 p\left(\overline{q}_c\right)\overline{h}_c - R\right\},
\end{aligned}$$

the first order conditions with respect to B^1, $\overline{q}_c$ and s are respectively given by:

$$\frac{\partial \mathcal{L}}{\partial B^1} = \left(\theta^1 + \gamma^1\right)\frac{\partial V^{1in}}{\partial B^1} - \gamma^1 \frac{\partial V^{1out}}{\partial B^1} - \lambda^{out}\frac{\partial \widehat{V}^{2out}}{\partial B^1} - \lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial B^1} + \frac{\partial \nu\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{dB^1} - \mu\pi^1 = 0, \tag{B25}$$

$$\frac{\partial \mathcal{L}}{\partial \overline{q}_c} = \left(\theta^1 + \gamma^1\right)\frac{\partial V^{1in}}{\partial \overline{q}_c} - \gamma^2 \frac{\partial V^{2in}}{\partial \overline{q}_c} - \lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial \overline{q}_c} + \frac{\partial \nu\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{d\overline{q}_c} - \mu\pi^1 \overline{h}_c p'\left(\overline{q}_c\right) = 0, \tag{B26}$$

$$\frac{\partial \mathcal{L}}{\partial \overline{h}_c} = \left(\theta^1 + \gamma^1\right)\frac{\partial V^{1in}}{\partial \overline{h}_c} - \gamma^2 \frac{\partial V^{2in}}{\partial \overline{h}_c} - \lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial \overline{h}_c} + \frac{\partial \nu\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{d\overline{h}_c} - \mu\pi^1 p\left(\overline{q}_c\right) = 0. \tag{B27}$$

Adding and subtracting $\lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial B^1}\frac{\partial V^{1in}}{\partial \overline{q}_c}/\frac{\partial V^{1in}}{\partial B^1}$ to (B26), one can rewrite it as:

$$\begin{aligned}
& \left[\left(\theta^1 + \gamma^1\right)\frac{\partial V^{1in}}{\partial B^1} - \lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial B^1}\right]\frac{\frac{\partial V^{1in}}{\partial \overline{q}_c}}{\frac{\partial V^{1in}}{\partial B^1}} + \lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial B^1}\frac{\frac{\partial V^{1in}}{\partial \overline{q}_c}}{\frac{\partial V^{1in}}{\partial B^1}} - \lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial \overline{q}_c} \\
& -\gamma^2\frac{\partial V^{2in}}{\partial \overline{q}_c} + \frac{\partial \nu\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{d\overline{q}_c} - \mu\pi^1\overline{h}_c p'\left(\overline{q}_c\right) \\
= \;& 0.
\end{aligned} \tag{B28}$$

From the first order condition (B25) one can derive an expression for $\left(\theta^1 + \gamma^1\right)\frac{\partial V^{1in}}{\partial B^1} - \lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial B^1}$ that can be substituted in (B28) to obtain:

$$\begin{aligned}
& \left[\gamma^1\frac{\partial V^{1out}}{\partial B^1} + \lambda^{out}\frac{\partial \widehat{V}^{2out}}{\partial B^1} - \frac{\partial \nu\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{dB^1} + \mu\pi^1\right]\frac{\frac{\partial V^{1in}}{\partial \overline{q}_c}}{\frac{\partial V^{1in}}{\partial B^1}} + \lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial B^1}\frac{\frac{\partial V^{1in}}{\partial \overline{q}_c}}{\frac{\partial V^{1in}}{\partial B^1}} \\
& -\lambda^{in}\frac{\partial \widehat{V}^{2in}}{\partial \overline{q}_c} - \gamma^2\frac{\partial V^{2in}}{\partial \overline{q}_c} + \frac{\partial \nu\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{d\overline{q}_c} - \mu\pi^1\overline{h}_c p'\left(\overline{q}_c\right) \\
= \;& 0.
\end{aligned} \tag{B29}$$

Denoting by $MRS^{1in}_{\overline{q}_c B}$ and $\widehat{MRS}^{2in}_{\overline{q}_c B}$ the marginal rate of substitution between $\overline{q}_c$ and B for, respectively, an opting-in type-1 household and an opting-in type-2 mimicker, one can rewrite (B23) as:

$$\begin{aligned}
MRS^{1in}_{\overline{q}_c B} &= \overline{h}_c p'(\overline{q}_c) + \frac{\lambda^{in}}{\mu\pi^1}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\left[\widehat{MRS}^{2in}_{\overline{q}_c B} - MRS^{1in}_{\overline{q}_c B}\right] - \frac{\lambda^{out}}{\mu\pi^1}\frac{\partial\widehat{V}^{2out}}{\partial B^1}MRS^{1in}_{\overline{q}_c B} \\
&\quad + \frac{\gamma^2}{\mu\pi^1}\frac{\partial V^{2in}}{\partial\overline{q}_c} - \frac{\gamma^1}{\mu\pi^1}\frac{\partial V^{1out}}{\partial B^1}MRS^{1in}_{\overline{q}_c B} \\
&\quad - \frac{1}{\mu\pi^1}\frac{\partial v\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{d\overline{q}_c} + \frac{1}{\mu\pi^1}\frac{\partial v\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{dB^1}MRS^{1in}_{\overline{q}_c B},
\end{aligned}$$

and thus, rearranging terms, one immediately obtains the result stated by (19).

Adding and subtracting $\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\frac{\partial V^{1in}}{\partial\overline{h}_c}/\frac{\partial V^{1in}}{\partial B^1}$ to (B27), one can rewrite it as:

$$\begin{aligned}
&\left[\left(\theta^1 + \gamma^1\right)\frac{\partial V^{1in}}{\partial B^1} - \lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\right]\frac{\frac{\partial V^{1in}}{\partial\overline{h}_c}}{\frac{\partial V^{1in}}{\partial B^1}} + \lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\frac{\frac{\partial V^{1in}}{\partial\overline{h}_c}}{\frac{\partial V^{1in}}{\partial B^1}} - \lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial\overline{h}_c} \\
&-\gamma^2\frac{\partial V^{2in}}{\partial\overline{h}_c} + \frac{\partial v\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{d\overline{h}_c} - \mu\pi^1 p(\overline{q}_c) \\
=\;& 0 \qquad \text{(B30)}
\end{aligned}$$

From the first order condition (B25) one can derive an expression for $\left(\theta^1 + \gamma^1\right)\frac{\partial V^{1in}}{\partial B^1} - \lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}$ that can be substituted in (B30) to obtain:

$$\begin{aligned}
&\left[\gamma^1\frac{\partial V^{1out}}{\partial B^1} + \lambda^{out}\frac{\partial\widehat{V}^{2out}}{\partial B^1} - \frac{\partial v\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{dB^1} + \mu\pi^1\right]\frac{\frac{\partial V^{1in}}{\partial\overline{h}_c}}{\frac{\partial V^{1in}}{\partial B^1}} \\
&+\lambda^{in}\frac{\partial\widehat{V}^{2in}}{\partial B^1}\left(\frac{\frac{\partial V^{1in}}{\partial\overline{h}_c}}{\frac{\partial V^{1in}}{\partial B^1}} - \frac{\frac{\partial\widehat{V}^{2in}}{\partial\overline{h}_c}}{\frac{\partial\widehat{V}^{2in}}{\partial B^1}}\right) - \gamma^2\frac{\partial V^{2in}}{\partial\overline{h}_c} + \frac{\partial v\left(q^1, q^2\right)}{\partial q^1}\frac{dq^1}{d\overline{h}_c} - \mu\pi^1 p(\overline{q}_c) \\
=\;& 0. \qquad \text{(B31)}
\end{aligned}$$

Denoting by $MRS^{1in}_{\overline{h}_c B}$ and $\widehat{MRS}^{2in}_{\overline{h}_c B}$ the marginal rate of substitution between $\overline{h}_c$ and B for, respectively, an opting-in type-1 household and an opting-in type-2 mimicker, one can rewrite

(B23) as:

$$
\begin{aligned}
& MRS^{1in}_{\overline{h}_c B} \\
= \; & p\left(\overline{q}_c\right) - \frac{1}{\mu\pi^1} \frac{\partial v\left(q^1, q^2\right)}{\partial q^1} \frac{dq^1}{d\overline{h}_c} + \frac{1}{\mu\pi^1} \frac{\partial v\left(q^1, q^2\right)}{\partial q^1} \frac{dq^1}{dB^1} MRS^{1in}_{\overline{h}_c B} \\
& + \frac{\gamma^2}{\mu\pi^1} \frac{\partial V^{2in}}{\partial \overline{h}_c} - \frac{\gamma^1}{\mu\pi^1} \frac{\partial V^{1out}}{\partial B^1} MRS^{1in}_{\overline{h}_c B} \\
& + \frac{\lambda^{in}}{\mu\pi^1} \frac{\partial \widehat{V}^{2in}}{\partial B^1} \left(\widehat{MRS}^{2in}_{\overline{h}_c B} - MRS^{1in}_{\overline{h}_c B} \right) - \frac{\lambda^{out}}{\mu\pi^1} \frac{\partial \widehat{V}^{2out}}{\partial B^1} MRS^{1in}_{\overline{h}_c B}.
\end{aligned}
$$

Given that we have observed that the optimal policy chosen by the government is such that $\overline{h}_c \leq h^{1in}_c$, we have that $MRS^{1in}_{\overline{h}_c B} = p\left(\overline{q}_c\right)$, which allows recovering the result stated by (20).

B Child Care Subsidies in the United States

In this section we describe the federal and state tax credits (for California) assuming a family with one child filing jointly.

At the federal level there are two tax credits. One is independent on whether a family had child care expenses or not. It is only based on the fact that the family has a dependent child. This tax credit (which is displayed in line 22 of the NBER taxsim "federal tax calculations") takes value 1.000 USD for all levels of family AGI (adjusted gross income) up to 110.000. Starting at an AGI of 110.000, it starts being phased out: for every 1.000 USD of AGI in excess of the 110.000 threshold, the value of the credit is reduced by 50 USD (for example, for an AGI=112.000 USD, the credit is equal to 1.000 – 2x50=900 USD). Thus, this credit goes to zero at AGI=130.000. The second federal tax credit is conditional on the family having incurred child care expenses (this credit is displayed in line 24 of the NBER taxsim "federal tax calculations"). This credit takes the following form:

$$
\beta^{FED}\left(Y^{AGI}\right) \cdot \min\left\{3.000, D, w^f L^f, w^m L^m\right\},
$$

where D denotes actual child care expenses for the family, 3.000 is a fixed amount, $w^f L^f$ is the earned income of the father, $w^m L^m$ is the earned income of the mother, and $\beta^{FED}\left(Y^{AGI}\right)$ takes value between 20% and 35% according to the decreasing schedule in table 16.

In addition we take into account the California child care tax credit which is a fraction of the second federal tax credit illustrated above (this State tax credit seems to have been refundable until 2011). (This credit is reported on line 38 of the NBER taxsim "State tax calculations".) Thus, the value of the State tax credit can be expressed as follows:

$$
\beta^{CAL}\left(Y^{AGI}\right) \cdot \beta^{FED}\left(Y^{AGI}\right) \cdot \min\left\{3.000, D, w^f L^f, w^m L^m\right\},
$$

Table 16: Federal and California tax credit schedule

Y^{AGI}	β^{FED}	Y^{AGI}	β^{FED}	Y^{AGI}	β^{CAL}
0 - 15000	35%	29,000- 31,000	27%	0 - 40,000	50%
15,000- 17,000	34%	31,000- 33,000	26%	40,000- 70,000	43%
17,000- 19,000	33%	33,000- 35,000	25%	70,000- 100,000	34%
19,000- 21,000	32 %	35,000- 37,000	24%	100,000-	0%
21,000- 23,000	31%	37,000- 39,000	23%		
23,000- 25,000	30%	39,000- 41,000	22%		
25,000- 27,000	29%	41,000- 43,000	21%		
27,000- 29,000	28%	43,000-	20 %		

where $\beta^{CAL}\left(Y^{AGI}\right)$ takes value between 0% and 50% according to the decreasing schedule in table 16.

C Optimal tax system with $\overline{v} = 0$

Table 17: Pure Nonlinear Income Tax Optimum

Allocation in households where the mother works

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$
43.879	42.866	0.199	0.437	0.238	0.165	0.397	1.166	22.738	0.11	0.000
56.353	51.056	0.279	0.416	0.214	0.172	0.415	1.474	29.597	0.113	0.066
69.357	59.076	0.315	0.4	0.203	0.177	0.42	1.811	36.645	0.114	0.136
91.963	72.346	0.368	0.4	0.187	0.177	0.436	2.305	47.66	0.114	0.125
148.942	103.999	0.455	0.425	0.16	0.17	0.47	3.426	73.948	0.112	0.000

Allocation in households where the mother does not work

y	c	L_m	L_f	h_m	h_f	h_c	q_c	q	D/y	$T'(M)$
35.487	34.732	0	0.434	0.303	0.167	0.33	1.165	20.324	0.113	0.162
39.715	36.75	0	0.423	0.306	0.17	0.323	1.428	24.836	0.121	0.261
46.849	40.228	0	0.434	0.309	0.167	0.324	1.703	29.855	0.122	0.283
52.648	42.433	0	0.419	0.313	0.173	0.314	2.061	35.857	0.128	0.388
86.692	58.846	0	0.542	0.315	0.135	0.35	2.969	54.579	0.125	0.000

Welfare Gain = 4.55% of GDP

Rows correspond to household types 1-5. Taxable income M and consumption c expressed in thousands of USD (2006 values).

References

Atkinson, A. B., and J. E. Stiglitz (1976). The Design of Tax Structure: Direct versus Indirect Taxation. *Journal of Public Economics* 6, 55-75.

Baker, M., Gruber, J. and K. Milligan (2008). Universal Child Care, Maternal Labor Supply, and Family Well-being. *Journal of Political Economy* 116 (4), 709-745.

Bastani, S., Blomquist, S. and L. Micheletto (2012). Child Care Subsidies, Quality and the Optimal Income Tax. In *Essays on the Economics of Income Taxation*, Economic Studies 129, Department of Economics, Uppsala University.

Bastani, S., Blomquist, S. and L. Micheletto (2013). The Welfare Gains of Age-related Optimal Income Taxation. *International Economic Review* 54 (4), 1219-1249.

Bastani, S., Blomquist, S. and J. Pirttilä (2015). How should commodities be taxed? A counter-argument to the recommendation in the Mirrlees Review. *Oxford Economic Papers* 67 (2), 455-478.

Becker G. and R. Posner (2005). Should Governments Subsidise Child Care and Work Leaves? Becker-Posner Blog, November 2005, www.becker.posner-blog.com/

Besley, T., Coate, S. (1991). Public Provision of Private Goods and the Redistribution of Income. *American Economic Review* 81 (4), 979-984.

Bick, A. (2016). The Quantitative Role of Child Care for Female Labor Force Participation and Fertility. *Journal of the European Economic Association* 14(3), 639-668.

Blackorby, C., Donaldson, D., (1988). Cash versus Kind, Self-selection and Efficient Transfers. *American Economic Review* 78 (4), 691-700.

Blau, D. and J. Carrie (2006). Pre-school, Child care, and After-school Care: Who's Minding the Kids? in *Handbook of the Economics of Education*, Elsevier, North Holland.

Blau, D. and P. K. Robins (1988). Child-care Costs and Family Labor Supply. *Review of Economics and Statistics* 70 (3), 374-381.

Blundell, R. and A. Shephard (2012). Employment, Hours of Work and the Optimal Taxation of Low-Income Families. *Review of Economic Studies* 79(2), 481-510.

Blomquist, S. and V. Christiansen (1995). Public Provision of Private goods as a Redistributive Device in an Optimum Income Tax Model. *Scandinavian Journal of Economics* 97 (4), 547-567.

Blomquist, S. and V. Christiansen (1998). Topping up or Opting out? The Optimal Design of Public Provision Schemes. *International Economic Review* 39, 399-411.

Blomquist, S., Christiansen, V. and L. Micheletto (2010). Public Provision of Private Goods and Nondistortionary Marginal Tax Rates. *American Economic Journal: Economic Policy* 2 (2), 1-27.

Blomquist, S., Christiansen, V. and L. Micheletto (2016). Public Provision of Private Goods, Self-selection and Income Tax Avoidance. *Scandinavian Journal of Economics* 118 (4), 666-692.

Boadway, R., Marchand, M., 1995. The Use of Public Expenditures for Redistributive Purposes. *Oxford Economic Papers* 47 (1), 45-59.

Carneiro, P. and J. Heckman (2003). Human Capital Policy. in J. Heckman and A. Krueger, Inequality in America: What role for Human Capital Policy? MIT Press.

Cogan, J. F. (1981). Fixed Costs and Labor Supply. *Econometrica* 49 (4), 945-963.

Corlett, W.J. and D.C. Hague (1953). Complementarity and the Excess Burden of Taxation. *Review of Economic Studies*, 21(1), 21-30.

Crawford, I., Keen, M., and S. Smith (2010). Value Added Taxes and Excises. In *Dimensions of Tax Design: The Mirrlees Review*, edited by Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., and J. Poterba. Oxford: Oxford University Press.

Cremer, H., Gahvari, F. (1997). In-kind Transfers, Self-selection and Optimal Tax Policy. *European Economic Review* 41 (1), 97-114.

Currie, J. (2006). The Invisible Safety Net: Protecting the Nation's Poor Children and Families. Princeton University Press.

Domeij, D. and P. Klein (2013). Should Day Care be Subsidized? *Review of Economic Studies* 80 (2), 568-595.

Farhi, E. and I. Werning (2010). Progressive Estate Taxation. *Quarterly Journal of Economics* 125, 635-673.

Gahvari, F. (1994). In-Kind Transfers, Cash Grants and Labor Supply. *Journal of Public Economics* 55, 495-504.

Gahvari, F. (1995). In-Kind versus Cash Transfers in the Presence of Distortionary Taxes. *Economic Inquiry* 33 (1), 45-53.

Gahvari, F. and E. Mattos (2007). Conditional Cash Transfers, Public Provision of Private Goods and Income Redistribution. *American Economic Review* 97, 491-502.

Guesnerie R. and K. Roberts (1984). Effective Policy Tools and Quantity Controls. *Econometrica* 52, 59-82.

Guner, N., Kaygusuz, R. and G. Ventura (2017). Child-Related Transfers, Household Labor Supply and Welfare, HCEO Working Paper No. 2017-001, The University of Chicago.

Gustafsson, S. and F. Stafford (1992). Child Care Subsidies and Labor Supply in Sweden. *Journal of Human Resources* 27 (1), 204-230.

Hausman, J. (1980). The Effects of Wages, Taxes and Fixed Costs on Women's Labor Force Participation. *Journal of Public Economics* 14, 61-194.

Havnes, T and M. Mogstad (2015). Is Universal Child Care Leveling the Playing Field? *Journal of Public Economics* 127, 100-114.

Havnes, T. and M. Mogstad (2011). No Child Left Behind: Subsidized Child Care and Children's Long Run Outcomes. *American Economic Journal: Economic Policy* 3, 97-129.

Heathcote, J., Storesletten, K. and G. L. Violante (2014). Consumption and Labor Supply with Partial Insurance: An Analytical Framework. *American Economic Review* 104(7), 2075-2126.

Heckman, J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science* 312 (5782), 1900-1902.

Heckman, J. and D. V. Masterov (2007). The Productivity Argument for Investing in Young Children. *Review of Agricultural Economics* 29(3), 446-493.

Ho, C. and N. Pavoni (2016). Efficient Child Care Subsidies. IGIER wp n. 572, Bocconi University.

Kaplow, L. (2009). On the Taxation of Private Transfers. *Tax Law Review* 63 (1), 159-188.

Kleven, H. J., Kreiner, C. T. and E. Saez (2009). The Optimal Income Taxation of Couples. *Econometrica* 77, 537-560.

Koehne, S. and D. Sachs (2017). Pareto Efficient Tax Breaks. Mimeo.

Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., and J. Poterba (eds.) (2010). *Dimensions of Tax Design: The Mirrlees Review*. Oxford: Oxford University Press.

Nichols, A. L. and R. J. Zeckhauser (1982). Targeting Transfers through Restrictions on Recipients. *American Economic Review, Papers and Proceedings* 72, 372-377.

OECD (2001). Starting Strong: Early Childhood Education and Care, Paris.

OECD (2006). Starting Strong II, Paris.

OECD (2011). Doing Better for Families.

OECD (2012). Starting Strong III, Paris.

OECD (2015). Starting Strong IV, Paris.

Powell, L. M. (2002). Joint Labor Supply and Childcare Choice Decisions of Married Mothers. *Journal of Human Resources* 37 (1), 106-128.

Ribar, D. C. (1995). A Structural Model of Child care and the Labor Supply of Married Women. *Journal of Labor Economics* 13 (3), 558-597.

Rosen S. (1996). Public Employment and the Welfare State in Sweden. *Journal of Economic Literature* 34, 729-740.

Smolensky, E. (2007) Children in the Vanguard of the U.S. Welfare State: A Review of Janet Currie's The Invisible Safety Net and Jane Waldfogel's What Children Need. *Journal of Economic Literature*, vol XLV, 1011-1023.

Stern, N. H. (1986). On the Specification of Labour Supply Functions. in Unemployment, Search and Labour Supply, ed. R. W. Blundell and I. Walker, 121-142. Cambridge: Cambridge University Press.

Tuomala, M. (2010). On Optimal Non-linear Income Taxation: Numerical Results Revisited. *International Tax and Public Finance* 17, 259-270.

Waldfogel, J. (2006) What Children Need, Harvard University Press.