Child Care Subsidies, Quality, and Optimal Income Taxation*

Spencer Bastani† Sören Blomquist‡ Luca Micheletto§

April 25, 2019

Abstract

We study the role of child care subsidies in a Mirrleesian optimal tax framework where parents choose both the quantity and quality of child care. Child care services not only enable parents to work, but also contribute to children's formation of human capital. We examine the conditions under which child care expenditures should be encouraged or discouraged by the tax system under different assumptions regarding the available policy instruments and what aspects of child care purchases that can be observed by the government at the individual level. Using a quantitative model calibrated to the US economy, we illustrate the possibility that child care expenditures should be taxed rather than subsidized, and discuss the merits of public provision schemes for child care.

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

JEL classification: H21; H41.

1 Introduction

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 enable both

*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.

†Department of Economics and Statistics, Linnaeus University; Linnaeus University Center for Integration and Discrimination Studies, Uppsala Center for Fiscal Studies, Uppsala Center for Labor Studies, CESifo. E-mail: spencer.bastani@lnu.se.

‡Uppsala Center for Fiscal Studies at the Department of Economics, Uppsala University. E-mail: soren.blomquist@nek.uu.se.

§Department of Law, University of Milan; Dondena Center for Research on Social Dynamics, Bocconi University; Uppsala Center for Fiscal Studies; CESifo. E-mail: luca.micheletto@unibocconi.it.

parents to work.[1] Another argument in favor of child care subsidies is that they represent a means to increase fertility (see, for example, OECD 2011). 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.[2]

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 against the deadweight losses of the taxes that are needed to finance the subsidies. The optimal tax literature has, however, pushed the opposite argument, namely, that subsidies to child care have the potential to increase the overall 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 subject to a more lenient tax treatment, and dates back to Corlett and Hague (1953).

In this paper we evaluate, both theoretically and quantitatively, the desirability of child care subsidies in a model where the quantity and quality of care that children receive, both at home (in terms of informal care from parents) and outside the home (at child care facilities), affects the children's human capital that enters into the utility function maximized by parents.

Our model economy consists of households that have the same (positive) number of children in child care ages but differ in terms of market ability and, possibly, also in terms of the production function for the children's human capital. The analysis is cast in a Mirrleesian optimal tax framework where the government pursues redistributive goals, but is constrained by asymmetric information. In particular, while the government knows the structure of households' preferences and the distribution of household types, it cannot tell "who is who". This implies that personalized lump-sum taxes and transfers are not feasible. The set of available tax instruments will then depend on which variables are publicly observable at the individual level and, possibly, on other constraints. In accordance with the bulk of the optimal tax literature, earned income will be assumed to be publicly observable at the individual level, so that a nonlinear earned income tax is an available policy instrument. Due to the asymmetric information problem, the government will in general need to impose distortions on agents' behavior in order to achieve redistributive goals. These distortions represent the efficiency costs that have to be incurred in order to

[1] See, for example, OECD (2006), as well as Blau and Robins (1988), Gustafsson and Stafford (1992), Ribar (1995) and Powell (2002).

[2] See, for example, Blau and Currie (2006), Currie (2006) and Waldfogel (2006), as well as Heckman (2006) and Heckman and Masterov (2007), who argue that high quality child care has the potential to help with the promotion of social skills, reduce rates of crime, teenage pregnancy, high school dropout rates, adverse health conditions and other social problems. In more recent work, Havnes and Mogstad (2011) find that subsidized child care has large positive effects on children's adult long-run outcomes and Havnes and Mogstad (2015) report that 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.

induce agents to truthfully reveal their types, and choose the bundle intended for them by the government, rather than behave as "mimickers" to reduce their tax burden. In such a setting, the desirability of supplementing income taxation with other instruments hinges crucially on the possibility to use them to make mimicking less attractive. As several previous contributions (that will be reviewed in Section 2) have pointed out, child care subsidies could be justified precisely on these grounds.

Compared to previous contributions in the optimal tax literature, the main distinguishing contribution of our paper is to emphasize the key importance of the quality dimension of child care (and of the human capital formation process more generally) in determining whether or not child care subsidies are useful as an instrument to achieve redistribution at lower efficiency costs.

In the first part of the paper, we theoretically analyze the role for child care subsidies in a simplified two-type setting, focusing on various government's problems that differ with respect to the economic variables that are assumed to be publicly observable at the individual level, and with respect to the policy instruments that are available to the government.

We start by characterizing the properties of a constrained-efficient allocation under the assumption that earned income, hours spent at a child care facility and quality of the chosen facility are all variables that are publicly observable at the individual level. We then analyze the properties of a constrained-efficient allocation under the assumption that only earned income and child care expenditures are publicly observable at the individual level. Under the same observational assumptions, we also analyze the government's problem under the constraint that, while earned income can be subject to a nonlinear income tax, child care expenditures can only be subsidized or taxed at a proportional rate. Finally, we consider the possibility that a nonlinear income tax is supplemented with child care subsidies that are administered through an "opting-out" public provision scheme.

The reason for considering these different cases is that it allows illustrating that the observational assumptions that one makes, as well as the assumptions regarding the policy instruments available to the government, can be relevant to determine whether child care expenditures should be encouraged or discouraged by the tax system.

In the second part of the paper, we construct a more general model that we calibrate to empirical wage distributions and time use patterns based on US data. Assuming that the government chooses optimally a nonlinear income tax on total household income, we evaluate the welfare effects of two alternative ways of subsidizing child care expenditures.

First, we consider offering child care subsidies at rates that are allowed to depend on both the household income and the labor force participation decision of the secondary earner. This is equivalent to allow a fraction (that is dependent both on the household income and the labor force participation decision of the secondary earner) of child care expenditures to be claimed as a refundable tax credit.

Second, we consider a public provision scheme where parents can choose between i) getting center-based child care services of a given fixed quality (chosen by the government) for as many hours as they want for free (in which case households are said to "opt-in"), and ii) choosing their preferred quality of center-based services but bearing the full cost (in which case households are said to "opt-out").[3]

Our numerical results, based on data for the US, indicate that subsidizing child care expenditures by means of refundable tax credits cannot be justified as a way to achieve redistribution at lower efficiency costs, at least not when the nonlinear income tax is optimally chosen. Intuitively, the reason is due to the fact that the tax-saving value of a refundable tax credit depends positively on the amount of child care expenditures. High-skilled mimickers, i.e. high-skilled individuals who reduce their labor supply in order to qualify for a more lenient tax treatment, can be expected to use formal child care for fewer hours than households being mimicked, i.e. actual low-skilled households. However, with quality of center-based care being a choice variable for households, high-skilled mimickers might still spend more on child care than low-skilled agents. This is indeed what happens in our simulations. In contrast, we find that administering a subsidy through an opting-out public provision scheme allows achieving redistribution at lower efficiency costs. Intuitively, the reason is that, by conditioning the subsidy on the acceptance of a quality that is set by the government, the tax-saving value of the subsidy, for opting-in households, only depends on the number of hours of center-based child care.

The rest of 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 theoretical model that is used to evaluate the desirability of subsidizing child care under different assumption regarding which economic variables are publicly observable at the individual level and which policy instruments are available to the government. In Section 4 we describe the model, calibrated on US data, that we employ to evaluate the social welfare effects of subsidizing child care; we also present the government's problem that we consider in the quantitative analysis as well as the computational approach used to solve the model. Section 5 provides the quantitative results of our numerical simulations. Finally, Section 6 offers some concluding remarks.

[3]The study of child care subsidies in the form of opting-out public provision schemes 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.

2 Related literature

While the emphasis on the importance of the quality dimension of child care (and the human capital formation process more generally) is a novelty of our analysis and suggests that child care subsidies might not necessarily be desirable, our paper relates to various previous contributions that have provided a role for child care subsidies as an efficiency-enhancing device.

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 and set optimally, commodity taxes are a redundant policy instrument when preferences are separable between leisure and other goods. If the separability condition is not satisfied, the theorem prescribes to 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. Viewing child care services as a primary example of services that are complements with labor supply, it would seem natural to argue that they should be subsidized or in any case be subject to a more lenient tax treatment (compared with other goods/services).[4] This is indeed the result obtained by Blomquist et al. (2010) in a model where i) the quality of center-based child care services does not enter into the utility function of agents, and ii) parents only use child care to cover for hours spent working in the market.

In another contribution, Domeij and Klein (2013) study how child care subsidies can help achieve 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 again, 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 more 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, as Domeij and Klein (2013), they do not consider a Mirrleesian optimal income tax setting, and even though the quantity of child care in their model 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.[5]

[4]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) for a recent discussion of this result.

[5]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 related paper is by Koehne and Sachs (2017), who study the benefits of providing tax deductions for household services in a Mirrleesian framework. However, whereas our focus is on describing properties of particular welfare optima in the context of child care, their work is focused on Pareto-improving reforms in the context of household services more generally. Hence, we view their work as complementary to ours.

Finally, the paper most closely related to ours is Ho and Pavoni (2016), who consider a static Mirrleesian setting and provide a rich set of results on the optimal manner in which to subsidize child care. Even though their setting is similar to ours, the two main differences are that they consider a different model of household decision-making and analyze the government's problem in a different way. The most important difference with respect to the first point, is that we endogenize the choice of quality of formal care. With respect to the second point, they compute the constrained efficient allocation on the basis of the informational frictions in the economy in a first step. Then, in a second step, they determine the discrepancies between marginal rates of substitution and marginal rates of transformation ("wedges") and proceed to discuss which policy instruments are needed to implement the constrained efficient allocation. In our paper, instead, we follow a complementary approach. In particular, while in the first part of the theoretical analysis we characterize the wedges prevailing at a constrained-efficient allocation, in the second part of the theoretical analysis, and in our quantitative analysis, we pre-specify the policy instruments and then analyze their qualitative and quantitative features. When we pre-specify the policy instruments, we consider a nonlinear income tax and either linear child care subsidies (linear tax credits) or a simple opting-out public provision scheme. A final difference between our paper and Ho and Pavoni (2016) is that we restrict ourselves to a two-type setting in our theoretical analysis, whereas they consider a more general theoretical framework with an arbitrary discrete number of household types.

3 A simplified theoretical model

In this section, we theoretically analyze child care subsidies using a simple two-type single-parent household model. This model will then be extended in several dimensions when we perform our quantitative analysis in section 4.

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 (in accordance with what is observed 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 consists of two types of single-parent households who differ in terms of market ability and, possibly, also in terms of the production function for the child's human capital.

We let households of type 1 be the low-ability households, and households of type 2 the high-ability ones, and denote by w^1 and w^2 the wage rate paid to, respectively, a high-ability agent and a low-ability agent (with $w^2 > w^1$). The total number of households is normalized to unity and the proportion of households of type 1 is denoted by π (so that $1-\pi$ is the proportion of households of type 2).

We also assume that children need always someone to take care of them. Care can either be provided by the parent (when he/she is not working in the market or is not engaged in other activities without the child) or by means of external child care services offered by centers which differ in quality.

Denoting by Θ the time endowment of a parent, his/her time constraint is given by:

$$L + h + \ell = \Theta, \tag{1}$$

where L represents the time devoted to work in the market, h denotes the time spent by parent with the child, and ℓ represents leisure time spent by the parent without the child.

Based on our assumption that children always need someone to take care of them, the time constraint for a child is:

$$h_c + h = \Theta, \tag{2}$$

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

To capture in a simple way the idea 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, we assume that human capital is built according to a function that depends on: i) the amount of time spent by a parent with the child; ii) the amount of time spent by the child at a child care facility; iii) the nurturing ability of the parent; iv) the quality of the child care facility chosen by the parent; v) the innate ability of the child. Formally, we let human capital be given by a function

$$\gamma f\left(\omega h, q_c h_c\right),$$

where $\omega\,(>0)$ denotes the nurturing ability of the parent, $q_c\,(\geq 0)$ denotes the quality of the child care facility chosen by the parent, and $\gamma\,(>0)$ is a scalar reflecting the innate ability of the child.

Parents derive utility from the consumption of a composite good denoted by c (treated as the *numéraire* of our economy), from leisure, and from the human capital of their child. In particular, households' preferences are represented by the following utility function:

$$U = u\left(c\right) + \gamma f\left(\omega h, q_c h_c\right) + v\left(\ell\right), \tag{3}$$

where we assume that $u\left(\cdot\right)$, $f\left(\cdot,\cdot\right)$ and $v\left(\cdot\right)$ are concave functions that are increasing in each argument.

Regarding the values of the parameters γ and ω, we will assume that they are weakly higher for parents of high market ability than for parents of low market ability.[6] Thus, we will focus on a setting where $w^2 > w^1$, $\gamma^2 \geq \gamma^1$ and $\omega^2 \geq \omega^1$.[7]

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 isoelastic function $p(q_c) = k(q_c)^\sigma$, with $k > 0$ and where we assume $\sigma \geq 1$.

Substituting the time constraints (1)-(2) into the utility function (3), we can write the problem solved by a parent under laissez faire as:

$$\max_{L,h,q_c} \quad u\left(wL - (\Theta - h)\, p(q_c)\right) + \gamma f\left(\omega h, (\Theta - h)\, q_c\right) + v\left(\Theta - L - h\right)$$

Using primes to denote derivatives and denoting by f_j' the derivative of the $f(\cdot,\cdot)$-function with respect to its j-th argument (with $j = 1, 2$), the first order conditions of the maximization problem above are given by:

$$L \quad : \qquad 1 - \frac{v'(\ell)}{wu'(c)} = 0 \tag{4}$$

$$h \quad : \qquad 1 - \frac{v'(\ell)}{p(q_c)\, u'(c) + (\omega f_1' - q_c f_2')\,\gamma} = 0 \tag{5}$$

$$q_c \quad : \qquad 1 - \frac{p'(q_c)\, u'(c)}{\gamma f_2'} = 0. \tag{6}$$

Below we will characterize the solutions to various different government's problems. In all these problems we will maintain the following two assumptions: i) the government aims at redistributing towards agents of type 1 (who are worse off under laissez-faire), and ii) the government knows the distribution of types in the population but does not know "who is who", so that first-best personalized lump-sum taxes and transfers are not feasible. On the other hand, the three problems that we will consider differ in terms of other observability assumptions and/or assumptions about the policy instrument available to the government. In particular, we will start by characterizing the properties of a constrained-efficient allocation under the assumption that earned income, hours spent by a child at a child care facility and quality of the chosen child care facility are all variables that are publicly observable at the individual level (case 1). We will then proceed to analyze the properties of a constrained-efficient allocation under the assumption that only earned income and child care expenditures are publicly observable at the individual level (case 2); in this case the tax function is allowed to be a general, non-separable

[6] We have previously said that the parameter γ captures the innate ability of the child in acquiring human capital. Alternatively, differences in γ across households could be interpreted as reflecting heterogeneity in preferences. Under this alternative interpretation, the human capital of the child carries a larger weight in the utility function of households characterized by a higher value for γ.

[7] The case $\gamma^2 > \gamma^1$ can be seen as reflecting the genetic inter-generational transmission of ability.

function of both earned income and child care expenditures. As a third case, we analyze the government's problem under the assumption that, while earned income can be subject to a nonlinear income tax, child care expenditures can only be subsidized or taxed at a proportional rate (case 3). This constraint can either be viewed as descending from the fact that child care expenditures are not publicly observable at the household level, or as capturing the idea that a proportional subsidy/tax is easier to administer and therefore allows to save on the administrative costs of taxes. Finally, we consider the possibility that a nonlinear income tax is supplemented with child care subsidies that are administered through an "opting-out" public provision scheme.

As will be clear at the end of our analysis, one reason for considering these various cases is that it allows highlighting that, depending on the observability assumptions that one makes and on the set of policy instruments available to the government, one can get opposite conclusions about the desirability to subsidize formal child care.

3.1 Case 1: Earned income, hours of child care, and quality observable

Assume that earned income, hours spent by a child at a child care facility and the quality of the chosen facility are all variables that are publicly observable at the individual level. Denoting earned income by y (with $y \equiv wL$), these observability assumptions imply that the tax function can be a general function $T = T(y, h_c, q_c)$ or equivalently, exploiting the time constraint (2), $T = T(y, h, q_c)$. A constrained efficient allocation can then be found as a solution to the following government's program:

$$\max_{c^1, y^1, h^1, q_c^1, c^2, y^2, h^2, q_c^2} \quad u\left(c^1\right) + \gamma^1 f\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right) + v\left(\Theta - \frac{y^1}{w^1} - h^1\right)$$

subject to

$$u\left(c^2\right) + \gamma^2 f\left(\omega^2 h^2, \left(\Theta - h^2\right) q_c^2\right) + v\left(\Theta - \frac{y^2}{w^2} - h^2\right) \geq \overline{V},$$

$$\begin{aligned} & u\left(c^2\right) + \gamma^2 f\left(\omega^2 h^2, \left(\Theta - h^2\right) q_c^2\right) + v\left(\Theta - \frac{y^2}{w^2} - h^2\right) \\ \geq \; & u\left(c^1\right) + \gamma^2 f\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right) + v\left(\Theta - \frac{y^1}{w^2} - h^1\right), \end{aligned}$$

$$\left[y^1 - c^1 - p\left(q_c^1\right)\left(\Theta - h^1\right)\right]\pi + \left[y^2 - c^2 - p\left(q_c^2\right)\left(\Theta - h^2\right)\right](1 - \pi) \geq \overline{R}.$$

In the maximization problem above the first constraint prescribes a minimum utility requirement for the high-skilled; the second constraint is a self-selection (incentive-compatibility) constraint requiring that high-skilled households have no incentive to mimic low-skilled households by choosing a bundle intended only for the latter. This constraint descends from the informational frictions in the economy (the fact that the government cannot directly observe the true type of an agent) and from the assumption that the socially desirable direction of redistribution is from type 2 to type 1. Finally, the last constraint represents the resource constraint of the economy (with $\overline{R}$ denoting an exogenous revenue requirement for the government).

The following Proposition characterizes the properties of the solution to the government's program.

Proposition 1. *No distortion should be imposed on the choices made by high-skilled households. For low-skilled households, y^1 should be downward distorted so that*

$$1 - v'\left(\Theta - \frac{y^1}{w^1} - h^1\right)/w^1 u'\left(c^1\right) > 0;$$

q_c^1 should be left undistorted if households only differ in terms of market ability ($\gamma^1 = \gamma^2 \equiv \gamma$ and $\omega^1 = \omega^2 \equiv \omega$); otherwise, if either $\omega^2 > \omega^1$ or $\gamma^2 > \gamma^1$ (or both $\omega^2 > \omega^1$ and $\gamma^2 > \gamma^1$), q_c^1 should be downward distorted so that

$$1 - p'\left(q_c^1\right) u'\left(c^1\right)/\gamma^1 f_2'\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right) > 0.$$

Finally, h_c^1 should be upward distorted (unless $y^1 = 0$, in which case h_c^1 should be left undistorted) if households only differ in terms of market ability;[8] *otherwise, if either $\omega^2 > \omega^1$ or $\gamma^2 > \gamma^1$ (or both $\omega^2 > \omega^1$ and $\gamma^2 > \gamma^1$), h_c^1 should be upward (resp.: downward) distorted when the following condition holds:*

$$\begin{aligned} & \gamma^2\left[\omega^2 f_1'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right) - q_c^1 f_2'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right)\right] - v'\left(\Theta - \frac{y^1}{w^2} - h^1\right) \\ > & \ (<) \\ & \gamma^1\left[\omega^1 f_1'\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right) - q_c^1 f_2'\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right)\right] - v'\left(\Theta - \frac{y^1}{w^1} - h^1\right). \end{aligned}$$

Proof See the Appendix. □

According to Proposition 1, all margins of choice for a high-skilled household should be left undistorted. Intuitively, since the government aims at redistributing from high- to low-skilled households, the latter have no incentive to behave as mimickers by falsely claiming to be of type 2 (i.e. high-skilled) in order to get the bundle (c^2, y^2, h^2, q_c^2). Therefore, distorting the choices of high-skilled households would entail efficiency losses

[8] That h_c^1 should be upward distorted is equivalent to say that h^1 should be downward distorted. Formally, it means that at a solution to the government's program the following condition is satisfied: $1 - v'\left(\Theta - \frac{y^1}{w^1} - h^1\right)/\left(p\left(q_c^1\right) u'\left(c^1\right) + \left[\omega f_1'\left(\omega h^1, \left(\Theta - h^1\right) q_c^1\right) - q_c^1 f_2'\left(\omega h^1, \left(\Theta - h^1\right) q_c^1\right)\right]\gamma\right) > 0$.

without any offsetting benefit in terms of mimicking-deterring effects.

Proposition 1 also states that incentive-compatibility considerations require to distort downwards the labor supplied in the market by low-skilled households, both in the case when market ability is the only dimension of heterogeneity among agents and in the case when high-skilled households are also more effective in building up human capital for their children ($\omega^2 > \omega^1$ or $\gamma^2 > \gamma^1$). The intuition for this result comes from the fact that, for a given amount of time devoted to children (h), the marginal disutility (in terms of foregone leisure) of earning an additional dollar is in both cases higher for a low-skilled than for a high-skilled household (formally: $-v'\left(\Theta - \frac{y^1}{w^1} - h^1\right)/w^1 < -v'\left(\Theta - \frac{y^1}{w^2} - h^1\right)/w^2$).

Regarding the quality of center-based care, q_c^1 should be left undistorted if households only differ in terms of market ability. This is because, for a given amount of time devoted to children (h), the marginal return of rasing q_c^1, in terms of higher human capital for the child, is the same for a low-skilled agent and for a high-skilled mimicker when $\gamma^1 = \gamma^2 = \gamma$ and $\omega^1 = \omega^2 = \omega$ (being equal to $\gamma f_2'(\omega h^1, (\Theta - h^1) q_c^1)$). However, if either $\omega^2 > \omega^1$ or $\gamma^2 > \gamma^1$ (or both $\omega^2 > \omega^1$ and $\gamma^2 > \gamma^1$), the marginal return of raising q_c^1 would be lower for a low-skilled household than for a high-skilled mimicker ($\gamma^1 f_2'(\omega^1 h^1, (\Theta - h^1) q_c^1) < \gamma^2 f_2'(\omega^2 h^1, (\Theta - h^1) q_c^1)$), in which case a downward distortion on q_c^1 would be optimal since it would make mimicking less attractive for high-skilled households.

Finally, according to Proposition 1, the direction of the optimal distortion on h_c^1 cannot be unambiguously determined unless households only differ in market ability. In such a case, h_c^1 should be upward distorted (or left undistorted if at an optimum $y^1 = 0$), which is equivalent to say (since from (2) it is $h_c = \Theta - h$) that h^1 should be downward distorted. The intuition for this result comes from the fact that, for a given quality of the child care facility, the marginal return of rasing h^1, in terms of higher human capital for the child, is the same for a low-skilled agent and for a high-skilled mimicker when $\gamma^1 = \gamma^2 \equiv \gamma$ and $\omega^1 = \omega^2 \equiv \omega$ (being equal to $\gamma\left[\omega f_1'(\omega h^1, (\Theta - h^1) q_c^1) - q_c^1 f_2'(\omega h^1, (\Theta - h^1) q_c^1)\right]$). At the same time, however, the marginal disutility (in terms of foregone leisure) of increasing h^1 is, for a given value of $y^1 > 0$, higher for a low-skilled than for a high-skilled household (formally, $-v'\left(\Theta - \frac{y^1}{w^1} - h^1\right) < -v'\left(\Theta - \frac{y^1}{w^2} - h^1\right)$).

If instead either $\omega^2 > \omega^1$ or $\gamma^2 > \gamma^1$ (or both $\omega^2 > \omega^1$ and $\gamma^2 > \gamma^1$), one cannot rule out the possibility that h_c^1 should be downward distorted (equivalently, that h^1 should be upward distorted). Intuitively, the reason is that, while it is still true that the marginal disutility (in terms of foregone leisure) of increasing h^1 (conditional on y^1) is higher for a low-skilled than for a high-skilled household, it might be the case that the marginal return of rasing h^1, in terms of higher human capital for the child, is larger for a low-skilled agent than for a high-skilled mimicker, i.e. it might be that

$$\begin{aligned} & \gamma^2 \left[\omega^2 f_1'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right) - q_c^1 f_2'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right)\right] \\ < \; & \gamma^1 \left[\omega^1 f_1'\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right) - q_c^1 f_2'\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right)\right]. \end{aligned} \tag{7}$$

Whereas the first circumstance favors distorting h_c^1 upwards, inequality (7) would call for distorting h_c^1 in the opposite direction.[9]

3.2 Case 2: Only earned income and child care expenditures publicly observable

Assume now that y is publicly observable at the individual level and that, albeit h_c and q_c are not publicly observable at the individual level, the government can observe how much each household spends on child care. Denoting child care expenditures by D (i.e. $D \equiv p(q_c) h_c = p(q_c)(\Theta - h)$), these observability assumptions imply that the tax function can be a general function of the form $T = T(y, D)$.

Since a tax function $T(y, D)$ associates to each pair (y, D) a corresponding amount of consumption c, the problem of choosing an optimal tax schedule can be equivalently stated as the problem of selecting two triplets (y^i, D^i, c^i), one for each type of household. Since from the time constraint (2) we have that $\Theta - h = h_c$, and given that $h_c = D/p(q_c)$, it follows that $\Theta - h = D/p(q_c)$. Therefore, we can equivalently express leisure ℓ as $D/p(q_c) - y/w$. Thus, given any triplet (y, D, c), an agent of type i solves

$$\max_{q_c^i} \quad u(c) + \gamma^i f\left(\left(\Theta - \frac{D}{p(q_c^i)}\right)\omega^i, \frac{D q_c^i}{p(q_c^i)}\right) + v\left(\frac{D}{p(q_c^i)} - \frac{y}{w^i}\right).$$

Denote the resulting "conditional" demand function by $\boldsymbol{q}_c^i(y, D, c)$ and notice that, due to the assumed separability between c and other arguments in the utility function, the conditional demand function is really only a function of y and D, i.e. $\boldsymbol{q}_c^i(y, D)$. Furthermore, denote the indirect utility function by

$$V^i(y, D, c) \equiv u(c) + \gamma^i f\left(\left(\Theta - \frac{D}{p(\boldsymbol{q}_c^i(y, D))}\right)\omega^i, \frac{D\boldsymbol{q}_c^i(y, D)}{p(\boldsymbol{q}_c^i(y, D))}\right) + v\left(\frac{D}{p(\boldsymbol{q}_c^i(y, D))} - \frac{y}{w^i}\right).$$

A constrained-efficient allocation can then be found as a solution to the following government's program:

$$\max_{y^1, D^1, c^1, y^2, D^2, c^2} V^1\left(y^1, D^1, c^1\right)$$

[9] Rewriting (7) as LHS > RHS where LHS $= \left[\gamma^2 f_2'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right) - \gamma^1 f_2'\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right)\right] q_c^1$ and RHS $= \gamma^2\omega^2 f_1'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right) - \gamma^1\omega^1 f_1'\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right)$, we can notice that LHS is strictly positive when either $\omega^2 > \omega^1$ (assuming $f_{12}'' > 0$) or $\gamma^2 > \gamma^1$. RHS, instead, is necessarily positive when $\gamma^2 > \gamma^1$ and $\omega^2 = \omega^1$, but has an ambiguous sign when $\omega^2 > \omega^1$. For instance, denoting by $\epsilon_{f_{1,\omega}'}$ the elasticity of f_1' with respect to ω (i.e. $\epsilon_{f_{1,\omega}'} \equiv \omega h f_{11}''/f_1'$), the sign of RHS is negative if $\gamma^2 = \gamma^1$ and $\epsilon_{f_{1,\omega}'} < -1$.

subject to

$$\begin{aligned} V^2\left(y^2, D^2, c^2\right) &\geq \overline{V}, \\ V^2\left(y^2, D^2, c^2\right) &\geq V^2\left(y^1, D^1, c^1\right), \\ \left(y^1 - c^1 - D^1\right)\pi + \left(y^2 - c^2 - D^2\right)(1-\pi) &\geq \overline{R}. \end{aligned}$$

Define $q_c^1 \equiv \boldsymbol{q}_c^1\left(y^1, D^1\right)$, $q_c^2 \equiv \boldsymbol{q}_c^2\left(y^2, D^2\right)$ and $\widehat{q}_c \equiv \boldsymbol{q}_c^2\left(y^1, D^1\right)$. The following Proposition characterizes the properties of the solution to the government's program.

Proposition 2. *No distortion should be imposed on the choices made by high-skilled households. For low-skilled households, y^1 should be distorted downwards (i.e., $1-v'\left(\ell^1\right)/w^1u'\left(c^1\right) > 0$) if households only differ in terms of market ability ($\gamma^1 = \gamma^2$ and $\omega^1 = \omega^2$); otherwise, if either $\omega^2 > \omega^1$ or $\gamma^2 > \gamma^1$ (or both $\omega^2 > \omega^1$ and $\gamma^2 > \gamma^1$), y^1 should be distorted downwards (resp.: upwards) when*

$$\frac{v'\left(\frac{D^1}{p\left(q_c^1\right)} - \frac{y^1}{w^1}\right)}{w^1} > (<) \frac{v'\left(\frac{D^1}{p\left(\widehat{q}_c\right)} - \frac{y^1}{w^2}\right)}{w^2}.$$

Finally, child care expenditures D^1 should be distorted downwards (resp.: upwards) if the following condition holds:

$$\frac{\gamma^2 f_2'\left(\left(\Theta - \frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2, \frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)}{p'\left(\widehat{q}_c\right)} > (<) \frac{\gamma^1 f_2'\left(\left(\Theta - \frac{D^1}{p\left(q_c^1\right)}\right)\omega^1, \frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)}{p'\left(q_c^1\right)}.$$

Proof See the Appendix. □

Compared to the results obtained in the previous subsection, the first thing to notice is that the changes in the informational assumptions do not impair the result that the choices made by high-skilled households are left undistorted. Since low-skilled households have no incentive to behave as mimickers by falsely claiming to be of type 2 (i.e. high-skilled), nothing can be gained by distorting the triplet (y^2, D^2, c^2) intended for the high-skilled.

The changes in the informational assumptions have instead an impact on the results pertaining to low-skilled households.

With respect to the optimal distortion on y^1, the result that the labor wedge ought to be positive (i.e. y^1 distorted downwards) does not necessarily hold. In particular, while it holds for sure if households differ only in market ability, it might be violated if high-skilled households are also more effective in building up human capital for their children ($\omega^2 > \omega^1$ or $\gamma^2 > \gamma^1$). The reason why it is no longer possible to unambiguously establish that the labor wedge is positive stems from the fact that, under the current informational assumptions, the government has a weaker control on h^1. The public observability of expenditures on child care by households implies that the government can control D^1. However, households maintain freedom of choice regarding the way to achieve a given level of child care expenditures, i.e. they can freely choose among all combinations of q_c^1 and h^1 satisfying $(\Theta - h^1)\, p\,(q_c^1) = D^1$. When households only differ in market ability, one can prove that, even though a high-skilled mimicker would choose a child care facility of higher quality than the one chosen by a low-skilled household ($\widehat{q}_c > q_c^1$) and would therefore spend more time with his child ($\widehat{h} > h^1$), the amount of leisure enjoyed by a mimicker would be higher than for a low-skilled. This in turn implies that, for given D^1, the marginal disutility (in terms of foregone leisure) of earning an additional dollar is higher for a low-skilled than for a high-skilled mimicker, and therefore that y^1 should be distorted downwards. However, when $\omega^2 > \omega^1$ or $\gamma^2 > \gamma^1$ (or both $\omega^2 > \omega^1$ and $\gamma^2 > \gamma^1$), one cannot rule out the possibility that the quality of the child care facility chosen by a mimicker is so much greater than the one chosen by a low-skilled, and therefore $\widehat{h}$ so much larger than h^1, that the amount of leisure enjoyed by a mimicker ends up being lower than for a low-skilled.[10]

Regarding the optimal distortion on D^1, remember that in the previous subsection we obtained the result that, when $\gamma^1 = \gamma^2$ and $\omega^1 = \omega^2$, q_c^1 should be left undistorted and h_c^1 should be upward distorted. Under the current informational assumptions, instead, the direction of the optimal distortion on D^1 cannot be in general unambiguously determined. This result indicates that the observability assumptions can play a key role in determining whether child care expenditures should be encouraged or discouraged by the tax system.

The following corollary of the second part of Proposition 2 illustrates that both a downward and an upward distortion on D^1 might be optimal when $\gamma^1 = \gamma^2$ and $\omega^1 = \omega^2$.[11]

Corollary 1. *i) Suppose $f_{12}'' > 0$, $p'' = 0$, $\gamma^1 \geq \gamma^2$ and $\omega^1 \geq \omega^2$; then, a downward distortion on D^1 is desirable.*

ii) Suppose $f_{12}'' = f_{22}'' = 0$, $p'' > 0$, $\gamma^1 = \gamma^2$ and $\omega^1 = \omega^2$; then, an upward distortion on D^1 is desirable.

Proof See the Appendix. □

[10]For instance, this would be the case if $w^2 = w^1$, $\omega^2 = \omega^1$, $\gamma^2 - \gamma^1 > 0$, and the price function is of the form $p\,(q_c) = kq_c$ (with $k > 0$). Under these assumptions, eq. (A31) in the Appendix ensures that $\widehat{q}_c > q_c^1$ and therefore, due to the assumption $w^2 = w^1$, $\widehat{\ell} = \frac{D^1}{p(\widehat{q}_c)} - \frac{y^1}{w^2} < \frac{D^1}{p(q_c^1)} - \frac{y^1}{w^1} = \ell^1$ and $v'\left(D^1/p\left(q_c^1\right) - y^1/w^1\right) \big/ w^1 < v'\left(D^1/p\left(\widehat{q}_c\right) - y^1/w^2\right) \big/ w^2$. The same would be true if, instead of $w^2 = w^1$ we had that $w^2 > w^1$ with the difference $w^2 - w^1$ sufficiently small. In fact, from eqs. (A29) and (A31) in the Appendix we know that we would still have $\widehat{q}_c > q_c^1$. Therefore, if the difference $w^2 - w^1$ were sufficiently small, it would still be that $v'\left(D^1/p\left(q_c^1\right) - y^1/w^1\right) \big/ w^1 < v'\left(D^1/p\left(\widehat{q}_c\right) - y^1/w^2\right) \big/ w^2$.

[11]With respect to the properties of the tax function $T\,(y, D)$ that allows implementing the constrained-efficient allocation, a downward (resp.: upward) distortion on D^1 implies that $\partial T\left(y^1, D^1\right)/\partial D^1 > 0$ (resp.: $\partial T\left(y^1, D^1\right)/\partial D^1 < 0$) at the $\left(y^1, D^1, c^1\right)$-bundle intended for low-skilled households.

The corollary above highlights the role played by the assumptions about the human capital production function and about the function relating the hourly price of center-based child care to its quality. The fact that both a downward and an upward distortion on D^1 might be optimal is interesting, especially when compared to the result that one would obtain in a model where agents only differ in market ability and there is no heterogeneity in the quality of the available child care facilities. Assuming that all child care facilities offer the same quality at a common hourly price, child care expenditures only depend on h_c, and one is immediately led to the conclusion that h_c^1 should be distorted upwards for mimicking-deterring purposes, i.e. child care expenditures should be subsidized at the margin for low-skilled households. This is the case on which previous contributions in the optimal tax literature have focused.

One strand of the literature (see, e.g., Blomquist et al., 2010, Bastani et al., 2015) simply assumed that agents demand a number of hours of center-based child care that is equal to their labor supply in the market (i.e., using our notation, $h_c = y/w$ for all agents), implicitly assuming that, for parents, time spent with the children is a perfect substitute for leisure time spent without the children. In this case child care services are simply viewed as something that needs to be purchased in order to enable agents to work.

Another strand of the literature (see, e.g., Ho and Pavoni, 2016) assumed instead that time spent by parents with the children is a perfect substitute for time spent working in the market; also in this case, however, the longer an agent works in the market, the less time he will spend with the children. Therefore, even though it is possible that agents demand child care hours in excess of their working hours (i.e., it is possible that $h_c > y/w$), it is still the case that center-based child care hours (and therefore child care expenditures) are increasing in the hours spent working in the market (y/w).

With child care expenditures being increasing in the number of hours spent working in the market, a high-skilled mimicker would always like to spend on child care less than a low-skilled agent (since for any given value for gross income y, a high-skilled mimicker needs to work fewer hours in the market). In accordance with the Atkinson-Stiglitz (1976) theorem, one would then benefit, for mimicking-deterring purposes, from supplementing a nonlinear income tax with other tax instruments aimed at encouraging the demand for child care services.[12] This is what drives the result, obtained in the previous literature, that child care expenditures should be subsidized (either linearly or nonlinearly).[13]

[12] According to the AS 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.

[13] In Blomquist et al. (2010) and Bastani et al. (2015), the assumption that $h_c = y/w$ for all agents leads to the conclusion that child care expenditures should be subsidized at a 100% rate. Ho and Pavoni (2016), instead, relaxing the assumption $h_c = y/w$ and allowing for nonlinear subsidies, find that child care subsidies should optimally follow a sliding scale. In a model with many consumption goods, one does not necessarily need to subsidize child care expenditures in order to encourage them; it would be

However, when agents have the possibility to pay a higher hourly price to get their children at a higher-quality facility, child care expenditures depend both on h_c and q_c. Thus, even if it were the case that a low-skilled favors having his child at a child care facility for more hours than a high-skilled mimicker, it would not necessarily follow that mimicking-deterring considerations require to subsidize at the margin child care expenditures by low-skilled agents. Contrary to previous findings in the optimal tax literature, downward distorting the child care expenditures of low-skilled households might then become desirable. In our model, as Corollary 1 shows, key elements to assess the direction of the optimal distortion on D^1 are given by the shape of the price function $p(q_c)$ and the properties of the human capital production function.

Having discussed the properties of a second-best optimum when y and D are publicly observable and the tax function is allowed to be a general function of the form $T = T(y, D)$, we can now move on to consider our next case.

3.3 Case 3: Earned income taxed non-linearly, child care expenditures subsidized or taxed at a proportional rate

Assume now that, whereas y can be subject to a nonlinear tax function $T(y)$ (that does not depend on D), child care expenditures can only be subsidized at a proportional (and income-independent) rate β (or taxed at a proportional rate if $\beta < 0$). Compared to the analysis presented in the previous subsection, to characterize the properties of a solution to the government's problem we will also need to take into account the constraint imposed by the assumed proportionality of child care subsidies.[14] For this purpose we will rely on an optimal revelation mechanism consisting of a set of type-specific before-tax incomes y^i and disposable incomes b^i (with $i = 1, 2$), and a proportional subsidy at rate β on child care expenditures. Thus, the mechanism assigns (β, b^i, y^i) to an agent who reports type i; the household then allocates b^i between child care expenditures and consumption of the composite consumption good c.[15]

Formally, given any triplet (β, b, y), a household of type i solves

$$\max_{q_c^i, h^i} \quad u\left(b - (1-\beta)\left(\Theta - h^i\right) p\left(q_c^i\right)\right) + \gamma^i f\left(\omega^i h^i, \left(\Theta - h^i\right) q_c^i\right) + v\left(\Theta - \frac{y}{w^i} - h^i\right).$$

enough to subject them to a more lenient tax treatment compared with other goods.

[14]As already remarked in the beginning of Section 3, this constraint can be interpreted in different ways. It could be viewed as descending from the fact that child care expenditures are not publicly observable at the household level, in which case the constraint would be a direct consequence of the informational frictions characterizing the economy. Alternatively, the constraint could be interpreted as capturing the idea that a proportional subsidy/tax is easier to administer and therefore allows saving on the administrative costs of taxes.

[15]Strictly speaking, this procedure does not characterize "allocations" as such; the optimization is over a mix of quantities and a tax rate (β). However, given β, utility maximizing households would choose the quantities themselves. We can thus think of the procedure as indirectly determining the final allocations.

Denote the resulting "conditional" demand functions by $\boldsymbol{q}_c^i(\beta, b, y)$ and $\boldsymbol{h}^i(\beta, b, y)$. Furthermore, denote the indirect utility function by

$$\begin{aligned} V^i(\beta, b, y) \equiv \; & u\left(b - (1-\beta)\left(\Theta - \boldsymbol{h}^i(\beta, b, y)\right) p\left(\boldsymbol{q}_c^i(\beta, b, y)\right)\right) \\ & + \gamma^i f\left(\omega^i \boldsymbol{h}^i(\beta, b, y), \left(\Theta - \boldsymbol{h}^i(\beta, b, y)\right) \boldsymbol{q}_c^i(\beta, b, y)\right) \\ & + v\left(\Theta - \frac{y}{w^i} - \boldsymbol{h}^i(\beta, b, y)\right). \end{aligned}$$

Define $q_c^1 \equiv \boldsymbol{q}_c^1(\beta, b^1, y^1)$, $q_c^2 \equiv \boldsymbol{q}_c^2(\beta, b^2, y^2)$, $\widehat{q}_c \equiv \boldsymbol{q}_c^2(\beta, b^1, y^1)$, $h^1 \equiv \boldsymbol{h}^1(\beta, b^1, y^1)$, $h^2 \equiv \boldsymbol{h}^2(\beta, b^2, y^2)$, $\widehat{h} \equiv \boldsymbol{h}^2(\beta, b^1, y^1)$. The government's problem can then be formally stated as:

$$\max_{y^1, b^1, y^2, b^2, \beta} V^1\left(\beta, b^1, y^1\right)$$

subject to

$$V^2\left(\beta, b^2, y^2\right) \geq \overline{V},$$

$$V^2\left(\beta, b^2, y^2\right) \geq V^2\left(\beta, b^1, y^1\right),$$

$$\left(y^1 - b^1 - \beta D^1\right)\pi + \left(y^2 - b^2 - \beta D^2\right)(1-\pi) \geq \overline{R},$$

and where $D^1 \equiv (\Theta - h^1)\, p(q_c^1)$ and $D^2 \equiv (\Theta - h^2)\, p(q_c^2)$.

Denote by $\widehat{D}$ the amount of child care expenditures for a high-skilled agent behaving as a mimicker (i.e. $\widehat{D} \equiv \left(\Theta - \widehat{h}\right) p(\widehat{q}_c)$) and define V_y^1, V_y^2, $\widehat{V}_y$, V_b^1, V_b^2 and $\widehat{V}_b$ as

$$\begin{aligned} V_y^1 &\equiv \partial V^1\left(\beta, b^1, y^1\right)/\partial y^1, \quad V_y^2 \equiv \partial V^2\left(\beta, b^2, y^2\right)/\partial y^2, \quad \widehat{V}_y \equiv \partial V^2\left(\beta, b^1, y^1\right)/\partial y^1, \\ V_b^1 &\equiv \partial V^1\left(\beta, b^1, y^1\right)/\partial b^1, \quad V_b^2 \equiv \partial V^2\left(\beta, b^2, y^2\right)/\partial b^2, \quad \widehat{V}_b \equiv \partial V^2\left(\beta, b^1, y^1\right)/\partial b^1. \end{aligned}$$

The following Proposition characterizes the properties of the solution to the government's program.

Proposition 3. *Define* $\left(\frac{dD^i}{dy^i}\right)_{dV^i=0}$ *as* $\left(\frac{dD^i}{dy^i}\right)_{dV^i=0} \equiv \frac{\partial D^i}{\partial y^i} - \frac{V_y^i}{V_b^i}\frac{\partial D^i}{\partial b^i}$. *For high-skilled households we have that:*

$$1 - \frac{v'\left(\Theta - \frac{y^2}{w^2} - h^2\right)}{w^2 u'(c^2)} = T'\left(y^2\right) = \left(\frac{dD^2}{dy^2}\right)_{dV^2=0} \beta. \tag{8}$$

Denote by λ *the Lagrange multiplier attached to the self-selection constraint and by* μ *the Lagrange multiplier attached to the resource constraint of the economy; for low-skilled households we have that:*

$$1 - \frac{v'\left(\Theta - \frac{y^1}{w^1} - h^1\right)}{w^1 u'(c^1)} = T'\left(y^1\right) = \left(\frac{dD^1}{dy^1}\right)_{dV^1=0} \beta + \frac{\lambda \widehat{V}_b}{\mu\pi}\left(\frac{\widehat{V}_y}{\widehat{V}_b} - \frac{V_y^1}{V_b^1}\right). \tag{9}$$

Finally, defining $\frac{\partial \widetilde{D}^i}{\partial \beta}$ *as* $\frac{\partial \widetilde{D}^i}{\partial \beta} \equiv \frac{\partial D^i}{\partial \beta} - D^i \frac{\partial D^i}{\partial b} > 0$, *the optimal proportional subsidy on child* *care is given by:*

$$\beta = \frac{\lambda \widehat{V}_b}{\mu} \frac{D^1 - \widehat{D}}{\pi \frac{\partial \widetilde{D}^1}{\partial \beta} + (1-\pi) \frac{\partial \widetilde{D}^2}{\partial \beta}};$$

therefore, $\beta > (<) \, 0$ *if* $D^1 > (<) \, \hat{D}$.

Proof See the Appendix. □

As shown in the first part of Proposition 3, the constraint imposed on the tax treatment of child care expenditures implies that, in contrast to what we obtained in the previous two subsections, the labor supply of high-skilled households is no longer left undistorted. Even though it is still the case that, in itself, nothing can be gained by distorting the behavior of high-skilled households, the fact that β is an income-independent proportional rate implies that, if it is optimal to set $\beta \neq 0$ to deter high-skilled households to behave as mimickers, the expenditure on child care by high-skilled households will necessarily be distorted. As a consequence, the marginal income tax rate faced by high-skilled households will deviate from zero in order to minimize the overall efficiency losses descending from distorting their behavior.

For the low-skilled households, the marginal income tax rate is given by the sum of a term that is the counterpart of the one determining the marginal income tax rate for high-skilled households, and a self-selection term that depends on the difference between the marginal rate of substitution between y and b for a low-skilled household and a high-skilled mimicker.

Finally, as shown in the last part of Proposition 3, a subsidy on child care expenditures is warranted if and only if a high-skilled behaving as a mimicker were to spend less on child care than a true low-skilled household. However, there is no guarantee that this is necessarily the case. The reason is the same that we discussed in the previous subsection. The difference, in this case, is that even when mimicking-deterring considerations call for distorting the child care expenditures of low-skilled households, the magnitude of the optimal subsidy rate (or tax rate) should also take into account that the subsidy (or tax) is going to apply to high-skilled households as well, distorting also their behavior and therefore producing additional efficiency losses (on top of those created by distorting the behavior of low-skilled households).[16]

[16] In the formula characterizing the optimal value for β in Proposition 3, the efficiency losses produced by setting $\beta \neq 0$ are captured by the sum $\pi \frac{\partial \widetilde{D}^1}{\partial \beta} + (1-\pi) \frac{\partial \widetilde{D}^2}{\partial \beta}$ appearing at the denominator of the expression on the right hand side. The term $\frac{\partial \widetilde{D}^i}{\partial \beta}$, which we have defined as $\frac{\partial D^i}{\partial \beta} - D^i \frac{\partial D^i}{\partial b}$, represents the change in the compensated gross expenditures on formal child care by households of type i, i.e. the change that occurs when a marginal increase in β is accompanied by a downward adjustment in b^i which leaves the utility of household i unchanged. Therefore, the term $\frac{\partial \widetilde{D}^i}{\partial \beta}$ captures the variation in D^i which is only due to substitution effects. If the utility function were linear in consumption, i.e. if $u'' = 0$, we would have that $\frac{\partial D^i}{\partial b} = 0$ and $\frac{\partial \widetilde{D}^i}{\partial \beta} = \frac{\partial D^i}{\partial \beta}$.

While the optimal sign of β cannot be in general unambiguously determined, the following Corollary shows that both $\beta < 0$ and $\beta > 0$ are possible outcomes.

Corollary 2. *i) Suppose* $f_{12}'' > 0$, $p'' = 0$, $\gamma^2 \geq \gamma^1$ *and* $\omega^2 \geq \omega^1$*; then,* $D^1 - \widehat{D} < 0$ *and it is optimal to levy a proportional tax on child care expenditures (*$\beta < 0$*).*

ii) Suppose $f_{12}'' = f_{22}'' = 0$, $p'' > 0$, $\gamma^1 = \gamma^2$ *and* $\omega^1 = \omega^2$*; then,* $D^1 - \widehat{D} > 0$ *and it is optimal to levy a proportional subsidy on child care expenditures (*$\beta > 0$*).*

Proof See the Appendix. □

Finally, Corollary 4 provides an example of an optimum where child care expenditures should be taxed and all agents face a positive marginal income tax rate.

Corollary 3. *Suppose* $u'' = 0$, $p'' = 0$, $f_{12}'' > 0$, $\gamma^1 = \gamma^2$ *and* $\omega^1 = \omega^2$*; then,* $\beta < 0$, $T'(y^2) > 0$, $T'(y^1) > 0$.

Proof See the Appendix. □

Having analyzed the properties of the solution to the government's program when the subsidy on child care expenditures is proportional and income-independent, it is straightforward to characterize the properties of an optimum when the proportional subsidy rate is allowed to be income-dependent. In such a case, the government would assign the triplet (β^i, b^i, y^i) to an agent who reports type i. Intuitively, the only difference with respect to the case considered in Proposition 3, is that the mimicking-deterring gains from distorting the child care expenditures of low-skilled households can now be reaped at lower efficiency costs. This is due to the fact that β^1 only applies to agents earning y^1. Since β^2 can be set independently of β^1, and given that there is no mimicking-deterring motive to distort the choices of high-skilled households, β^2 will be optimally set equal to zero, implying (replacing β in (8) with $\beta^2 = 0$) that high-skilled face a zero marginal income tax rate. The following Corollary summarizes the results for this case.

Corollary 4. *Assume that child care expenditures can be subsidized (or taxed) at a proportional but income-dependent rate. Then,*

i) all margins of choice for high-skilled households are left undistorted;

ii) low-skilled households face a marginal income tax rate that is still given by (9), but with β^1 *replacing* β*, where* β^1 *is given by the following expression:*

$$\beta^1 = \frac{\lambda \widehat{V}_b}{\mu \pi \frac{\partial \widetilde{D}^1}{\partial \beta}} \left(D^1 - \widehat{D} \right).$$

Proof See the Appendix. □

3.4 An opting-out public provision scheme

The analysis carried out in the previous subsections shows that the case for subsidizing center-based child care expenditures is significantly weakened once one takes into account that the market provides households with an array of choices regarding the quality of child care services. Ideally, one would like to have at disposal tax instruments that can distinguish between the share of child care expenditure that is due to a specific quality of the chosen facility, and the share that is due to the number of hours that a child spends at a facility. If a subsidy is "blind" to this difference, and in contrast to previous findings in the literature, a tax on child care might well be optimal.

However, even when hours spent by a child at a child care facility and the quality of the chosen facility are not publicly observable at the household level, there might be a way to come close to replicate the constrained-efficient allocation characterized in Proposition 1. In particular, suppose that a nonlinear income tax $T(y)$ is supplemented with an opting out public provision scheme that works as follows. The government provides center-based child care services of quality $\overline{q}_c$ at a given hourly user charge. Households can decide to "opt-in" and use the publicly provided services for as many hours as they need. Alternatively, if households are not happy with the quality provided by the government, they can "opt-out" and choose their preferred quality of center-based services paying the full market price.

To illustrate how child care subsidies administered through a public provision scheme may be desirable in cases when a *tax* on child care expenditures would otherwise be optimal, assume that $p'' = 0$, $w^2 > w^1$, $\gamma^2 = \gamma^1 \equiv \gamma$ and $\omega^2 = \omega^1 \equiv \omega$. Under these assumptions we know, from Corollary 2 part i), that when a proportional tax/subsidy on child care expenditures is used alongside an optimal nonlinear income tax, child care expenditures should be taxed; we also know that, at a pure income tax optimum (i.e. when a nonlinear income tax is the only instrument at disposal to the government), $\widehat{h} > h^1$ and $\widehat{q}_c > q_c$ (see eqs. (A59)-(A60) in the Appendix).

Suppose then that the government has the possibility to supplement a nonlinear income tax with an income-dependent opting out public provision system such that households earning y^1 have the possibility to choose between i) getting publicly provided child care services of a given quality $\overline{q}_c$ (chosen by the government) at an hourly user charge $(1-\beta)\,p\left(\overline{q}_c\right)$, and ii) choosing their preferred quality of center-based child care services but bearing the full cost.[17]

Formally, the problem solved by a household of type i who chooses to opt-in is:

$$\max_{h^i} \quad u\left(b^1 - (1-\beta)\left(\Theta - h^i\right)p\left(\overline{q}_c\right)\right) + \gamma f\left(\omega h^i, \left(\Theta - h^i\right)\overline{q}_c\right) + v\left(\Theta - \frac{y^1}{w^i} - h^i\right),$$

[17]The assumption that only households earning a given amount of income are eligible to opt-in is made to keep the analysis as simple as possible. We will come back to this later.

with associated first order condition

$$(1-\beta)\,p\,(\overline{q}_c)\,u' + \gamma\,(\omega f_1' - \overline{q}_c f_2') = v'. \tag{10}$$

Denote by $\boldsymbol{h}^i\,(\overline{q}_c, \beta, b^1, y^1)$ the "conditional" demand function for a household of type i who opts-in and denote the corresponding indirect utility function by

$$\begin{aligned} V^i\left(\overline{q}_c \beta, b^1, y^1\right) \;\equiv\;& u\left(b^1 - (1-\beta)\left(\Theta - \boldsymbol{h}^i\left(\overline{q}_c, \beta, b^1, y^1\right)\right) p\,(\overline{q}_c)\right) \\ &+\gamma f\left(\omega \boldsymbol{h}^i\left(\overline{q}_c, \beta, b^1, y^1\right), \left(\Theta - \boldsymbol{h}^i\left(\overline{q}_c, \beta, b^1, y^1\right)\right) \overline{q}_c\right) \\ &+v\left(\Theta - \frac{y^1}{w^i} - \boldsymbol{h}^i\left(\overline{q}_c, \beta, b^1, y^1\right)\right). \end{aligned}$$

For a household of type i choosing a bundle (b, y) and opting-out, the optimization problem is instead the following:

$$\max_{h^i, q_c^i}\; u\left(b - \left(\Theta - h^i\right) p\left(q_c^i\right)\right) + \gamma f\left(\omega h^i, \left(\Theta - h^i\right) q_c^i\right) + v\left(\Theta - \frac{y}{w^i} - h^i\right),$$

with associated first order conditions

$$\begin{aligned} p\left(q_c^i\right) u' + \gamma\,(\omega f_1' - \overline{q}_c f_2') &= v', \\ p'\left(q_c^i\right) u' &= \gamma f_2'. \end{aligned}$$

Denote by $\boldsymbol{h}^i\,(b, y)$ and $\boldsymbol{q}_c^i\,(b, y)$ the "conditional" demand functions for a household of type i who chooses the bundle (b, y) and opts-out. The corresponding indirect utility function is given by:

$$\begin{aligned} V^i\,(b, y) \;\equiv\;& u\left(b - \left(\Theta - \boldsymbol{h}^i\,(b, y)\right) p\left(\boldsymbol{q}_c^i\,(b, y)\right)\right) \\ &+\gamma f\left(\omega \boldsymbol{h}^i\,(b, y), \left(\Theta - \boldsymbol{h}^i\,(b, y)\right) \boldsymbol{q}_c^i\,(b, y)\right) + v\left(\Theta - \frac{y}{w^i} - \boldsymbol{h}^i\,(b, y)\right). \end{aligned}$$

Finally, define $h^{1,in} \equiv \boldsymbol{h}^1\,(\overline{q}_c, \beta, b^1, y^1)$, $h^2 \equiv \boldsymbol{h}^2\,(b^2, y^2)$, $q_c^2 \equiv \boldsymbol{q}_c^2\,(b^2, y^2)$, $\widehat{h}^{in} \equiv \boldsymbol{h}^2\,(\overline{q}_c, \beta, b^1, y^1)$, $\widehat{h} \equiv \boldsymbol{h}^2\,(b^1, y^1)$ and $\widehat{q}_c \equiv \boldsymbol{q}_c^2\,(b^1, y^1)$. The government's problem can be formally stated as:

$$\max_{y^1, b^1, y^2, b^2, \overline{q}_c, \beta} V^1\left(\overline{q}_c, \beta, b^1, y^1\right)$$

subject to

$$V^2\left(b^2, y^2\right) \geq \overline{V},$$

$$V^2\left(b^2, y^2\right) \geq V^2\left(b^1, y^1\right),$$

$$V^2\left(b^2, y^2\right) \geq V^2\left(\overline{q}_c, \beta, b^1, y^1\right)$$

$$\left[y^1 - b^1 - \beta p\left(\overline{q}_c\right)\left(\Theta - h^{1,in}\right)\right]\pi + \left(y^2 - b^2\right)(1-\pi) \geq \overline{R}.$$

In the problem above, there are two self-selection constraints since there are now two different mimicking strategies available to high-skilled households. One mimicking strategy is for them to choose the bundle (b^1, y^1) and opt-out; the other available strategy is for them to choose the bundle (b^1, y^1) and opt-out. The second and the third constraint in the government's problem above are the two self-selection constraints that jointly ensure incentive-compatibility.[18]

Rather than characterizing the optimal values for β and $\overline{q}_c$, suppose to start from an initial equilibrium that is a pure income tax optimum (an optimum where only a nonlinear income tax is used). Denote by q_c^{1*} the quality optimally chosen by low-skilled households at this equilibrium. Notice that, by setting $\overline{q}_c = q_c^{1*}$ and $\beta = 0$ in the above government's problem, one would replicate the pure income tax optimum.[19] Even though the proposed public provision scheme with $\overline{q}_c = q_c^{1*}$ and $\beta = 0$ is welfare-neutral, and therefore as such is really a redundant instrument, it is a useful starting point to shed light on how an opting-out public provision scheme may be welfare-enhancing.

To illustrate this point, our strategy will be the following. We will construct a policy reform that is welfare-neutral for low-skilled households and that at the same time replicates the pattern of distortions for h_c^1 and q_c^1 that are deemed desirable according to the constrained optimum characterized in Proposition 1. We will then evaluate the effects of the envisaged reform on the constraints faced by the government in maximizing the utility of low-skilled households. To simplify the presentation, below we will characterize the structure of our policy reform for the case when utility is quasi-linear in consumption so that $u''(c) = 0$.[20]

Remember that according to the results stated in Proposition 1, when agents only differ in market ability ($\gamma^2 = \gamma^1 \equiv \gamma$ and $\omega^2 = \omega^1 \equiv \omega$), a constrained-efficient allocation requires to impose an upward distortion on h_c^1 and to leave q_c^1 undistorted.

Thus, starting from our initial equilibrium where $\overline{q}_c = q_c^{1*}$ and $\beta = 0$, we will consider a reform that marginally raises β by $d\beta > 0$, while at the same time adjusting $\overline{q}_c$ and b^1

[18] Relaxing the assumption that eligibility for public provision is conditional on income would require to also consider in the government's problem how a policy change affects the opting-in/opting-out choice for high-skilled not behaving as mimickers. This however would not affect the main qualitative insights of our analysis.

[19] One would obtain the same values for y^1, b^1, y^2 and b^2 that prevail at a pure income tax optimum. Low-skilled households would be indifferent between i) choosing $\left(b^1, y^1\right)$ and opt-in, and ii) choosing $\left(b^1, y^1\right)$ and opt-out. The maximum utility that high-skilled households could obtain by behaving as mimickers is the same as under a pure income tax optimum given that they can always choose $\left(b^1, y^1\right)$ and opt-out.

[20] We have performed the same analysis also for the case when $u'' < 0$. The qualitative results remain unscathed but the expression become much more complicated.

in such a way that the following conditions are satisfied:

$$dV^1\left(\overline{q}_c,\beta,b^1,y^1\right)=\left(\Theta-h^{1,in}\right)p\left(\overline{q}_c\right)u'd\beta+u'db^1+ \\ +\left[\gamma f_2'-\left(1-\beta\right)p'\left(\overline{q}_c\right)u'\right]\left(\Theta-h^{1,in}\right)d\overline{q}_c=0, \quad (11)$$

$$\gamma\left(\omega f_{12}''-\overline{q}_c f_{22}''\right)\left(\frac{dh^{1,in}}{d\beta}+\frac{dh^{1,in}}{d\overline{q}_c}d\overline{q}_c+\frac{dh^{1,in}}{db^1}db^1\right)+\left(\Theta-h^{1,in}\right)\gamma f_{22}''d\overline{q}_c=0, \quad (12)$$

where all the various derivatives of the f-function are evaluated at $(\omega h^{1,in},(\Theta-h^{1,in})\,\overline{q}_c)$.

Condition (11) postulates that the reform is welfare-neutral for low-skilled households; condition (12) postulates that the overall effect of the reform is to leave the quality of child care services at an undistorted value for low-skilled households who opt-in.[21]

Exploiting the fact that at the initial equilibrium, $\beta=0$ and $\overline{q}=q_c^{1*}$, so that $\gamma f_2'-(1-\beta)\,p'\left(\overline{q}_c\right)u'=0$, we show in appendix A.8 that we must have that:

$$d\overline{q}_c = \frac{\left(\omega f_{12}''-\overline{q}_c f_{22}''\right)p\left(\overline{q}_c\right)u'}{\gamma\left(\Theta-h^{1,in}\right)(\omega)^2\left[\left(f_{12}''\right)^2-f_{11}''f_{22}''\right]-\left(\Theta-h^{1,in}\right)v''f_{22}''}d\beta<0, \quad (13)$$

$$db^1 = -\left(\Theta-h^{1,in}\right)p\left(\overline{q}_c\right)d\beta<0. \quad (14)$$

Notice that, since low-skilled households were indifferent between opting-in and opting-out at the pre-reform equilibrium, the proposed reform, while keeping their utility unchanged, makes them strictly prefer to opt-in. Notice also that, since low-skilled households prefer to opt-in, they will choose $h^{1,in}$ in accordance to (10), which implies (given that $\beta>0$ at the post-reform equilibrium) that their choice for hours of center-based child care is upward distorted.

Consider now the effects of the proposed reform, which is welfare-neutral by construction for low-skilled households, on the constraints faced by the government. Since the reform did not change y^2 and b^2, the reform is clearly also welfare-neutral for high-skilled households who do not behave as mimickers. Regarding the effect on the government's budget constraint, denoting by μ the Lagrange multiplier attached to this constraint, we have that the effect is given by:

$$\mu\pi\left\{-db^1-p\left(\overline{q}_c\right)\left(\Theta-h^{1,in}\right)d\beta-\beta p'\left(\overline{q}_c\right)\left(\Theta-h^{1,in}\right)d\overline{q}_c+\beta p\left(\overline{q}_c\right)dh^{1,in}\right\}.$$

However, taking into account that β was equal to zero at the pre-reform equilibrium and

[21] Condition (12) is obtained by totally differentiating the no-distortion condition

$$\gamma f_2'\left(\omega h^{1,in},\left(\Theta-h^{1,in}\right)\overline{q}_c\right)-p'\left(\overline{q}_c\right)u'\left(b^1-(1-\beta)\left(\Theta-h^{1,in}\right)p\left(\overline{q}_c\right)\right)=0,$$

and taking into account that we are assuming $u''=0$. Notice that at the initial equilibrium, where $\overline{q}_c=q_c^{1*}$ and $\beta=0$, the condition above is satisfied. This is obvious since the initial equilibrium replicates the pure income tax optimum. At a pure income tax optimum, the individual choices for h_c and q_c satisfy the no-distortion conditions (5) and (6).

substituting $-\left(\Theta - h^{1,in}\right) p\left(\bar{q}_c\right) d\beta$ for db^1, the expression above is equal to zero. Thus, the proposed reform is also budget-neutral for the government.

What is left to assess is then the impact of the reform on high-skilled households who were to behave as mimickers. In appendix A.9, we show that the proposed reform is mimicking-deterring in the sense that it mitigates both the self-selection constraints that appear in the government's problem.

Summarizing, what we have shown is that, starting from a pure income-tax optimum, it is possible to implement an opting-out public provision scheme that i) replicates the pattern of distortions for h_c^1 and q_c^1 that are deemed desirable according to a constrained optimum; ii) is welfare-neutral for all households not behaving as mimickers; iii) is budget-neutral for the government; iv) lowers the utility of high-skilled mimickers under both available mimicking strategies. Hence, a reform along the suggested lines opens the possibility for the government to achieve a Pareto-improvement upon the initial equilibrium.

4 Quantitative Model

We now proceed with the quantitative analysis. We first describe the quantitative model and our calibration which is carried out taking into account the current US tax system and existing child care subsidies. In section 4.7, we turn to our optimal tax analysis where we let the government optimize the nonlinear income tax schedule and the child care subsidy scheme.

4.1 General setting

The quantitative model is an extension of the simplified model considered in section 3. Most importantly we extend the setting to more than two types and consider two-earner couples consisting of mothers and fathers. We make 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 $i \in \{1, \ldots, N\}$ which we refer to as the *household* skill type (where a higher index corresponds to a higher ability).[22] Moreover, in line with previous literature (e.g., Cogan 1981, Blundell and Shephard 2012), we also assume that in each household, there is a fixed cost associated with the mother's labor force participation.[23] More specifically, we

[22]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.

[23]These costs enable the model to replicate empirically relevant shares of non-working households, and can be viewed, for example, as representing the psychological costs associated with leaving a child in the care of a non-parent.

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. 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 be summarized into a unidimensional parameter, the type-space is still bi-dimensional by virtue of the heterogeneity in the fixed costs of work.

Since our emphasis in the quantitative analysis is 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, the wage pair (w_m^i, w_f^i), where the subscript m refers to "mother" and the subscript f refers to "father", represents households where the mother belongs to the i:th skill category in the wage distribution for mothers and w_f^i is the average wage of all fathers matched with type-i mothers. This implies that the skill level of the household is described by the mothers' relative position (rank) in the wage distribution of mothers.[24] The procedure to compute the wages based on actual data is described in detail in section 4.3 below. In our simulations we consider $N = 5$ skill types and a continuous distribution of cost types (approximated by $\Psi = 1000$).[25]

4.2 Household decision problem

Let L_k, h_k and ℓ_k denote, respectively, the labor supply, domestic care, and leisure of spouse j (for $k = m, f$) and let w_k and ω_k denote, respectively, the wage rate and nurturing ability of spouse k. Moreover, denote by h_c and q_c the quantity and quality of formal care, and denote by q the human capital of the child, with q given by:

$$q = \gamma^i f(\omega_m^i h_m, \omega_f^i h_f, q_c h_c).$$

The general formulation of the problem solved by each household $(i, j) \in \Theta$ in the presence of existing taxes and child care subsidies can be described as follows:

$$\max_{c, L_m, L_f, h_m, h_f, \ell_m, \ell_f, h_c, q_c} \left\{ u(c) + g(q) + v(\ell_m) + v(\ell_f) - \mathbf{1}[L_m > 0] \cdot \chi^{ij} \right\}$$

subject to the household budget constraint

$$c = w_f^i L_f + w_m^i L_m - T^{US}(w_f^i L_f + w_m^i L_m) - CE(D, w_f^i L_f, w_m^i L_m), \tag{15}$$

[24] 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.

[25] Computational considerations prevent us from expanding the model beyond five skill types. However, taking into account that for each skill type we have both two-earner households and one-earner households, the model features 10 household types.

and the following time constraints (for, respectively, the mother, the father and the child):

$$L_m + h_m + \ell_m = 1, \tag{16}$$

$$L_f + h_f + \ell_f = 1, \tag{17}$$

$$h_c + h_m + h_f = 1. \tag{18}$$

In the above formulation we have normalized to one the time endowment for each member of the household. Moreover, $\mathbf{1}[L_m > 0]$ is an indicator function that takes the value 1 when both spouses are working, 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 $D \equiv p(q_c)\, h_c$ and the income of both spouses, $w_f^i L_f$ and $w_m^i L_m$.

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 a sample of households filing jointly with small children (below the age of 6), and assuming *zero* child care expenditure, we find $\tau = 0.164$ and $\lambda = 1.31$.

In constructing the functional form for the CE function we take into account the CTC (Child Tax Credit) and the CDCTC (Child and Dependent Care Tax Credit), which are federal tax credits.[26] In addition, we model the state tax credit that applies in California (and which is a function of the CDCTC), and the subsidies that are offered through the CCDF (Child Care and Development Fund).[27] 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.

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 replace the CE function with optimized child care subsidies.

[26]The CTC is not, strictly speaking, a child care subsidy since there is no requirement that paid child care be used. Thus, it is actually a child subsidy. Eligibility for the CTC is not conditional on being employed and its amount is (weakly) decreasing in the household AGI (Adjusted Gross Income). The CDCTC is a non-refundable tax credit available to families with children aged under 13 and covers part of child care expenses. Being employed is a requirement that must be fulfilled to be eligible for the CDCTC.

[27]US states are offered a block grant from the federal government in the form of the CCDF. The purpose of the CCDF is to increase the availability, affordability, and quality of child care services. In appendix B we describe in more detail the rules governing the various subsidy programs that we model in our analysis.

4.3 Data on market wage rates, participation and hours of work

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 and then take averages 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 level 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.[28] The wage distributions for mothers and fathers are approximated using the deciles of the predicted wage distributions.

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.[29] In addition, some variation in annual hours of work stems from the number of weeks worked during a year. For simplicity, we multiply weekly earnings by 48 to get a measure of annual earnings. Since we do not want to overfit our model to potentially noisy wage data, our calibration procedure targets average working hours for mothers and fathers.

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

[28]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.

[29]There is also a variable in the CPS called "hours worked last week" that 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).

Table 1: Hourly wage rates (2006 USD), weekly hours of work, and labor force participation rates (LFP) for mothers and fathers in our data.

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

4.4 Functional forms

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

$$U = c^{\alpha} + q^{\beta} - \frac{1}{\ell_m} - \frac{1}{\ell_f} - \mathbf{1}_{[L_m>0]} \cdot \chi^{ij}. \tag{19}$$

In the above specification we have employed functional forms that are suitable for numerical computation. The household derives utility from private consumption and the child's human capital through the concave functions c^{α} and q^{β} where $0 < \alpha < 1$ and $0 < \beta < 1$. In addition, the household derives utility from spousal leisure ℓ_m and ℓ_f through the increasing and concave function $v(x) = -1/x$ in order to produce desirable labor supply behavior.[30] The human capital of the child is assumed to depend on the overall quality of the child care arrangement, q, that is given by:

$$q = \gamma^i f(\omega_m^i h_m, \omega_f^i h_f, q_c h_c) = \gamma^i \left[\omega_m^i h_m^{\rho} + \omega_f^i h_f^{\rho} + q_c h_c^{\rho}\right]^{\frac{1}{\rho}}. \tag{20}$$

Eq. (20) is a three-input CES-production function where the relative importance of the production factors is endogenous due to the fact that q_c is a choice variable. There are few estimates available on the elasticity of substitution between different modes of child care, and this is clearly an area where more research is needed. Since our model is already quite rich, we adopt the normalization $\rho = 0.5$. We furthermore assume that home care productivities are related to market productivity by assuming that $\omega_m^i = \omega_f^i = \nu \frac{w_m^i + w_f^i}{2}$ where $\nu > 0$ is a type independent scale parameter used to ensure that the scale of ω and q_c are comparable inside the production function. Notice that since mothers and fathers in a given household have the same ω, mothers generally have a comparative advantage in household work due to their lower market wage rate (except in the highest skilled household where the market wage rate is higher for mothers). Finally, the production

[30] The form of the leisure term in the utility function has previously been employed in the optimal tax context by, for example, Tuomala (2010).

function in (20) features the parameter γ^i, which is set equal to $\frac{w_m^i}{\overline{w}_m}$ (where $\overline{w}_m$ is the average wage rate of mothers). This parameter captures differences in the innate abilities of children, due for instance to some degree of genetic transmission of ability within households.[31]

The hourly price of child care is a function of the quality chosen by agents and is described by the iso-elastic function $p(q_c) = kq_c^{\sigma}$. This is equivalent to the log-log relationship between the hourly price of child care and child care quality used by Blau and Mocan (2002). We further assume that $k > 0$ and $\sigma > 1$, implying a convex price function, in line with the relationship described by Havnes and Mogstad (2015).

4.5 Calibration

The parameters of the model are jointly optimized in order to minimize a loss function with the purpose of matching the following empirical targets: (i) average hours of work for mothers and fathers, (ii) female labor force participation rates, (iii) the profile of child care expenditure as a fraction of total household income across the household skill distribution (iv) the profile of the average hourly cost of child care across the household skill distribution. We describe how we match each of these targets in the paragraphs below.

Hours of work 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. the 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. Given our assumption that 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, a father who works 40 hours a week will have $L_f = 0.33$. Consistent with the data reported in table 1, we calibrate the average labor supply of mothers in the model to approximately be equal to $\frac{34.4 \times 48}{5840} \approx 0.28$ and $\frac{44 \times 48}{5840} \approx 0.36$ for fathers.

Hours of domestic care Table 2 describes how married women with children in ages 0-6 divide their time. 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

[31] The particular specification we use is motivated by the fact that we let the skill type of the household be determined by the mother's wage rate. In appendix D, we provide a robustness check where we let the parameter be defined based on the average household wage rate.

$4\times 7+48\times 2 = 124$ holidays/weekend days. Using the value for non-employed mothers to apply during weekends, we use $\frac{1.9\times 240}{5840}+\frac{4.2\times 124}{5840}\approx 0.17$ as the target for the fraction of the time endowment full-time employed mothers spend in domestic care in our calibration.[32]

Table 2: 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)

Labor force participation 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. More specifically, the distributions of fixed costs associated with mothers' labor force participation are chosen so that the model, under the benchmark US tax system, matches the household-specific motherly employment rates in table 1.[33] For this purpose, we have proceeded in the following way. For each skill type i, we compute the fixed cost that would make a mother 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 working mothers 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 mothers of type i, and a higher fixed cost for the remaining part of the population. We achieve this by assuming that the fixed costs are

[32]For part-time employed mothers the corresponding number is $\frac{2.7\times 240}{5840}+\frac{4.2\times 124}{5840}\approx 0.20$, and for non-working mothers, the target is $\frac{4.2\times 52\times 7}{5840}\approx 0.26$. Our calibration is, by and large, consistent with these numbers as well. The labor supply for a full-time working mother in the ATUS is, according to table 2, equal to 6.7 hours per working day. This amounts to $6.7\times 5 = 33.5$ hours per week which is in line with the average labor supply reported in the CPS (see table 1).

[33]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.

given by the power function[34]

$$\chi_j = a_0(j-1)^{a_1} + a_2, \quad j = 1, \ldots, \Psi$$

where we estimate the parameters a_0, a_1, and a_2 using nonlinear least squares on the data points $\{0, 0\} \cup \{\Psi z^i, \bar{\chi}^i\}_{i=1}^N$.[35] Intuitively, the relationship is monotonically increasing.

Pattern of child care expenditure We want our model to produce realistic patterns of child care expenditure as a fraction of family income across the household skill distribution. For this purpose, we have examined two sources of data. The Annual Social and Economic Supplement of the Current Population Survey (CPS ASEC) as well as the Survey of Income and Program Participation (SIPP). The important empirical feature of both these data sources is that the ratio of child care expenditure to family income is higher for low income families as compared to high income families in households where the mother works. In our calibration, we target the numbers provided by Laughlin (2013) who finds, using the 2008 SIPP, that child care expenditure as a fraction of family income ranges between around 25% at the bottom to around 15% at the top of the income distribution.[36]

Hourly price of child care To pin down the parameters of our price function, we target empirical patterns of the hourly price of child care, $p(q_c)$, across the household skill distribution for two-earner couples. We use the figures reported by Whitehurst (2018, page 9) who finds that the hourly price of child care ranges between around 3 USD to around 9 USD across the household income distribution.

4.6 The calibrated economy

A summary of the calibrated parameters in the model and their calibrated values is presented in table 3. This table also contains a summary of the empirical targets that are used in our calibration as well as the corresponding variables in our model output (see also table 4 below).[37]

[34] A power distribution for the fixed costs of work has previously been used by Kleven et al. (2009).

[35] The estimated parameters are $a_0 = 4.109e - 11$, $a_1 = 3.518$ and $a_2 = 0.005028$. Notice that the slope of the fixed cost distribution is related to the concept of "participation elasticity" emphasized in the public finance literature. Our approach implies heterogeneous participation elasticities depending on the skill-specific employment level, where the relationship is established through structural assumptions and a calibration procedure.

[36] Calculated based on Laughlin (2013), table 6. Another calculation is Herbst (2015), who finds that the corresponding figures range between around 17% at the bottom and 8% at the top (Herbst 2015, table 7).

[37] Notice that the parameters are jointly chosen to match the empirical targets using a model-fitting procedure.

Table 3: Summary of parameter values and calibration targets

Parameter values

Parameter	Value	Interpretation
α	0.52	Curvature utility private consumption
β	0.49	Curvature utility human capital
k	0.60	Scale parameter price function
σ	4.10	Curvature parameter price function
ν	0.10	Scale parameter household production

Calibration targets

Variable	Target	Model	Description
L_m	0.28	0.27	Avg. hours of work mothers
L_f	0.36	0.38	Avg. hours of work fathers
h_m	0.17	0.15	Avg. hours of maternal care
D/y	0.25–0.15	0.25-0.15	Expenditure share profile
$p(q_c)$	3–9 USD	2.7-8.9	Hourly cost profile

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

The column $1 - CE/D$ in table 4 shows the fraction of child care expenditure that is paid by the government. 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 zero for all one-earner households. For two-earner households the effective subsidy ranges between 51% and 10%, and is monotonically decreasing in the skill type of households.

The column T/y reports the average income tax rate paid by the various households, which ranges between 5% (for one-earner households of type 1) and 18% (for two-earner households of type 5). The column q shows that the human capital of children is increasing in the skill type of an household. It is lowest in households of type 1 where the mother does not work, and highest in households of type 5 where the mother does not work. The same pattern characterizes the overall quality of the child care arrangement.[38]

[38] According to (20) the human capital q depends also on γ^i, which captures the effect of the innate ability of a child and is increasing in the skill type of an household. Interpreting q/γ^i as a measure of the overall quality of the child care arrangement chosen by a household, we have that its profile is consistent with the profile of q: both among one-earner couples and two-earner couples it is increasing in the skill type of an household; moreover, it is lowest in households of type 1 where the mother does not work and highest in households of type 5 where the mother does not work.

Table 4: Benchmark allocation (calibrated economy)

Allocation in households where the mother works

i	y	c	L_m	L_f	h_m	$\frac{D}{y}$	q	q_c	$p(q_c)$	$\frac{T}{y}$	$T'(y)$	$1-\frac{CE}{D}$	U
1	44.50	36.28	0.18	0.45	0.08	0.25	0.81	1.45	2.72	0.06	0.22	0.51	1.50
2	55.46	42.83	0.25	0.42	0.11	0.22	1.48	1.51	3.29	0.10	0.25	0.40	2.34
3	66.00	49.07	0.28	0.40	0.14	0.20	2.42	1.58	3.94	0.12	0.27	0.31	3.17
4	77.88	55.91	0.30	0.35	0.18	0.18	4.16	1.68	5.09	0.15	0.29	0.22	4.24
5	98.04	67.69	0.32	0.26	0.24	0.15	9.66	1.93	8.89	0.18	0.31	0.10	6.43

Allocation in households where the mother does not work

i	y	c	L_m	L_f	h_m	$\frac{D}{y}$	q	q_c	$p(q_c)$	$\frac{T}{y}$	$T'(y)$	$1-\frac{CE}{D}$	U
1	41.24	34.74	0.00	0.50	0.14	0.11	0.76	1.18	1.18	0.05	0.21	0	1.36
2	47.09	37.93	0.00	0.50	0.21	0.12	1.50	1.31	1.81	0.07	0.23	0	2.10
3	53.45	41.58	0.00	0.49	0.27	0.13	2.53	1.43	2.59	0.09	0.24	0	2.86
4	60.13	45.32	0.00	0.48	0.34	0.14	4.44	1.58	3.93	0.11	0.26	0	3.87
5	70.13	51.08	0.00	0.44	0.43	0.14	10.43	1.89	8.21	0.13	0.28	0	5.98

In the table, y and c denote annual household income and consumption, respectively, expressed in thousands of USD (2006 values). Moreover, L_j denotes labor supply, CE net child care expenditure, D gross child care expenditure, T income tax liability, q overall quality of child care arrangement, q_c quality level of paid care arrangement, U household utility. Finally, $1 - CE/D$ is the implicit child care subsidy rate in the current US tax system.

4.7 Optimal tax systems

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. We describe the computational challenges and our computational approach in appendix C.

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. 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.[39] This allows us to identify, for each skill group, a unique marginal worker that is indifferent between working and not-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.[40] 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 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. 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.

We assume that the government maximizes the sum of individuals' utilities, subject to a concave transformation $W(\cdot)$ (reflecting society's taste for redistribution). Thus, the

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

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

welfare gain of subsidizing child care will be measured in terms of the effectiveness of raising social welfare. Following Brewer et al. (2010) we focus on a logarithmic transformation of individual utility.[41] We further assume that two distinct nonlinear income tax schedules apply to one-earner and two-earner households.[42]

In the absence of any kind of subsidies to child care expenditures (i.e. in the case of a pure income tax optimum), the government's problem can be described as follows:

$$\max_{\mathbf{P}} \Omega(\mathbf{P}) \tag{21}$$

$$\Omega(\mathbf{P}) = \max_{\{(y_1^i,b_1^i),(y_0^i,b_0^i)\}_{i=1}^N} \sum_{i=1}^{N} \left(\sum_{j>\mathbf{P}_i} \pi^{ij} W\left(V_0^i(y_0^i,b_0^i)\right) + \sum_{j\leq\mathbf{P}_i} \pi^{ij} W\left(V_1^i(y_1^i,b_1^i) - \chi_{ij}\right)\right) \tag{22}$$

$$\text{subject to:} \tag{23}$$

$$V^{ij}\left(y_0^i,b_0^i,y_1^i,b_1^i\right) \geq \tilde{V}^{ij}\left(y_0^{i-1},b_0^{i-1},y_1^{i-1},b_1^{i-1}\right), \forall i \in \{2,\ldots,N\}, \forall j \tag{24}$$

$$V^{ij}\left(y_0^i,b_0^i,y_1^i,b_1^i\right) = \begin{cases} V_0^i(y_0^i,b_0^i) & \text{if } j > \mathbf{P}_i \\ V_1^i(y_1^i,b_1^i) - \chi^{ij} & \text{if } j \leq \mathbf{P}_i \end{cases} \tag{25}$$

$$\tilde{V}^{ij}(y_0^{i-1},b_0^{i-1},y_1^{i-1},b_1^{i-1}) = \max\left\{V_0^i(y_0^{i-1},b_0^{i-1}), V_1^i(y_1^{i-1},b_1^{i-1}) - \chi^{ij}\right\} \tag{26}$$

$$V_0^i(y_0^i,b_0^i) > V_1^i(y_1^i,b_1^i) - \chi^{ij} \quad \forall i, \quad j > \mathbf{P}_i \tag{27}$$

$$V_0^i(y_0^i,b_0^i) < V_1^i(y_1^i,b_1^i) - \chi^{ij} \quad \forall i, \quad j \leq \mathbf{P}_i \tag{28}$$

$$\sum_i \left(\sum_{j\leq\mathbf{P}_i} \pi^{ij}(y_0^i - b_0^i) + \sum_{j>\mathbf{P}_i} \pi^{ij}(y_1^i - b_1^i)\right) \geq \overline{R} \tag{29}$$

$$V_0^i(y,b) = \max_{q_c,h_c,h_m} u^i\left(b - p(q_c)h_c, h_m, 0, \frac{y}{w_f^i}, h_c, q_c\right), \forall i \tag{30}$$

$$V_1^i(y,b) = \max_{q_c,h_c,h_m,L_m} u^i\left(b - p(q_c)h_c, h_m, L_m, \frac{y - w_m^i L_m}{w_f^i}, h_c, q_c\right), \forall i. \tag{31}$$

The first thing to notice is that the government's problem features three levels of optimization. Eq. (21) 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. 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 (24). These constraints ensure that each household prefers the bundle assigned to it rather than the bundle intended for the adjacent lower skilled household.

[41] In a previous version of the paper we have analyzed other social welfare functions, obtaining similar qualitative results as in the present paper.

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

Equations (25) and (26) 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 (26) 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 $i - 1$.[43] Inequalities (27) and (28) are individual rationality constraints that ensure that the labor force participation decisions prescribed in the $\mathbf{P}$ vector are actually the ones maximizing household utility. Constraint (29) 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 $\overline{R}$.[44] The last two equations define the indirect utilities for households where the mother does not work (eq. 30) and the indirect utilities (gross of the fixed cost of work) for households where the mother works (eq. 31). The computation of these two indirect utilities for each type-i household represents the lower level optimization problem.[45]

For the upper layer, that is responsible for finding the optimal participation vector $\mathbf{P}$, we use a global optimization heuristic that relies on a combination of coarse searches over the full parameter space and local searches around the best coarse point.[46] For the middle and lower layers, i.e. the bi-level optimization problem, we rely on an efficient implementation in C++, interfacing the latest version of the state-of-the-art solver for nonlinear constrained optimization problems KNITRO.

5 Quantitative Results

In our quantitative analysis we consider three cases. In each of these cases the government chooses optimally a nonlinear tax on household income. In the first case, the nonlinear income tax is the only policy instrument. In the second case, the government can also subsidize child care expenditures at a proportional rate, which is allowed to depend on both the mother's employment status and household income. Finally, in the third case, we allow for a simple opting-out public provision scheme where the quality of the publicly provided care is optimally chosen by the government and free of charge for opting-in

[43] 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.

[44] In our numerical simulations the revenue requirement $\overline{R}$ is always set equal to the fiscal surplus that arises in our US benchmark economy (described in subsection 4.6).

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

[46] 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 certain 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.

households (i.e., opting-in households can get free of charge as many hours of formal care as they want).[47]

The results for the case where the government optimizes a nonlinear income tax and there are no subsidies to child care are shown in table 5.

Table 5: Optimal nonlinear income tax

Allocation in households where the mother works

i	y	c	L_m	L_f	h_m	$\frac{D}{y}$	q	q_c	$p(q_c)$	$\frac{T}{y}$	$T'(y)$	β	U
1	47.22	45.09	0.21	0.47	0.08	0.12	0.93	1.22	1.34	0.05	0.19	-	2.07
2	58.04	50.24	0.27	0.44	0.11	0.13	2.14	1.34	1.99	0.13	0.26	-	3.17
3	79.85	62.84	0.36	0.46	0.11	0.13	4	1.46	2.87	0.21	0.14	-	4.43
4	94.8	71.35	0.37	0.42	0.15	0.14	8.58	1.61	4.22	0.25	0.17	-	6.41
5	141.02	100.76	0.44	0.39	0.18	0.14	26.5	1.9	8.31	0.29	0	-	11.13

Allocation in households where the mother does not work

i	y	c	L_m	L_f	h_m	$\frac{D}{y}$	q	q_c	$p(q_c)$	$\frac{T}{y}$	$T'(y)$	β	U
1	38.05	37.77	0	0.47	0.14	0.12	0.98	1.2	1.26	0.01	0.28	-	1.93
2	45.17	40.98	0	0.48	0.21	0.14	2.27	1.33	1.91	0.09	0.25	-	2.92
3	48.85	42.38	0	0.45	0.27	0.14	4.49	1.44	2.67	0.13	0.33	-	4.1
4	61.53	49.14	0	0.49	0.34	0.14	9.2	1.6	4.11	0.2	0.2	-	5.87
5	83.03	62.95	0	0.52	0.43	0.14	27.98	1.92	8.62	0.24	0	-	10.29

Household taxable income y and consumption c expressed in thousands of USD (2006 values).

In table 6 we show the results where the government employs income-dependent child care subsidies on top of an optimal nonlinear income tax. Glancing at the column labeled by β, it can immediately be seen that child care subsidies would be suboptimal. Instead, child care expenditure should be taxed ($\beta < 0$). The implied taxes on child care appear to be quite substantial (and, conditional on the mother's employment status, decreasing in household income), even though the implied welfare gain of these child care taxes is close to zero.[48] The result that there should be no subsidies, at least when levied as income-dependent proportional subsidies, challenges previous findings in the optimal tax literature advocating child care subsidies as an instrument to relax the incentive constraints faced by the government in designing an optimal nonlinear income tax.

[47] In these three cases, the optimal labor force participation rates for mothers in households of type 1 through 5 are 50%, 60%, 65%, 75%, and 85%, respectively.

[48] The welfare gain is calculated by computing the minimum amount of extra revenue that needs to be injected into pure optimal nonlinear income tax solution in order to reach the social welfare of the optimal income tax solution with child care subsidies. We then divide this amount of extra revenue by the aggregate pre-tax income in the pure optimal income tax solution to get a welfare measure expressed as a fraction of aggregate output.

Table 6: Income-dependent subsidy

Allocation in households where the mother works

i	y	c	L_m	L_f	h_m	$\frac{D}{y}$	q	q_c	$p(q_c)$	$\frac{T}{y}$	$T'(y)$	β	U
1	46.79	46.1	0.2	0.47	0.08	0.1	0.87	1.16	1.1	0.01	0.19	-0.18	2.08
2	58.2	51.74	0.27	0.44	0.11	0.11	2.04	1.28	1.65	0.11	0.25	-0.17	3.17
3	79.65	63.29	0.36	0.46	0.11	0.12	3.94	1.44	2.67	0.21	0.14	-0.06	4.42
4	94.89	71.71	0.37	0.42	0.15	0.13	8.51	1.59	4.06	0.24	0.17	-0.03	6.4
5	141.12	100.61	0.44	0.39	0.18	0.14	26.48	1.9	8.3	0.29	-0	-0	11.12

Allocation in households where the mother does not work

i	y	c	L_m	L_f	h_m	$\frac{D}{y}$	q	q_c	$p(q_c)$	$\frac{T}{y}$	$T'(y)$	β	U
1	38.28	39.96	0	0.47	0.14	0.08	0.88	1.09	0.86	-0.04	0.25	-0.38	1.95
2	45.18	42.08	0	0.48	0.21	0.11	2.18	1.27	1.61	0.07	0.24	-0.16	2.92
3	49.14	44.16	0	0.46	0.28	0.11	4.3	1.36	2.12	0.1	0.3	-0.22	4.09
4	61.65	49.54	0	0.49	0.34	0.14	9.12	1.58	3.9	0.2	0.19	-0.05	5.86
5	83.07	62.8	0	0.52	0.43	0.14	27.98	1.91	8.61	0.24	0	0	10.28

Household taxable income y and consumption c expressed in thousands of USD (2006 values).

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 is likely to spend 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.[49]

The discussion above suggests that in order for there to be a role for child care subsidies as an instrument to achieve redistribution at a lower efficiency cost, one needs a policy instrument that allows to control the quality of the child care services that are being subsidized. This would be for instance the case with an opting-out public provision scheme. Under an opting-out public provision scheme, 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

[49] 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.

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).

The quantitative results for this case are displayed in table 7. In that table, it can be seen that the government selects a quality level for the publicly provided child care such that all but the highest skilled households opt in. The quality is set at 1.42, which is a quality level close to the one chosen by median households in the absence of child care subsidies. Moreover, it is worth noticing that the bottom three household types are better off under the public provision scheme, whereas the top two households are worse off. The welfare gain of the adopting a public provision scheme for child care alongside an optimal nonlinear income tax is equal 1.43% of GDP. Thus, the public provision scheme appears to significantly improve upon the pure optimal nonlinear income tax solution.

Table 7: Public Provision

Allocation in households where the mother works

i	y	c	L_m	L_f	h_m	$\frac{D}{y}$	q	q_c	$p(q_c)$	$\frac{T}{y}$	$T'(y)$	Opt in	U
1	47.3	44.16	0.22	0.47	0.05	0.23	1.1	1.42	2.54	0.07	0.23	Yes	2.24
2	58.15	50.85	0.28	0.43	0.08	0.17	2.16	1.42	2.54	0.13	0.3	Yes	3.32
3	80.23	65.63	0.37	0.46	0.09	0.12	3.65	1.42	2.54	0.18	0.17	Yes	4.5
4	96.91	77.55	0.38	0.43	0.13	0.09	7.23	1.42	2.54	0.2	0.19	Yes	6.35
5	144.79	95.47	0.45	0.4	0.17	0.13	25.97	1.87	7.87	0.21	-0	No	10.75

Allocation in households where the mother does not work

i	y	c	L_m	L_f	h_m	$\frac{D}{y}$	q	q_c	$p(q_c)$	$\frac{T}{y}$	$T'(y)$	Opt in	U
1	38.4	36.84	0	0.47	0.09	0.27	1.15	1.42	2.54	0.04	0.29	Yes	2.11
2	45	40.99	0	0.48	0.14	0.2	2.29	1.42	2.54	0.09	0.28	Yes	3.07
3	48.13	42.86	0	0.45	0.2	0.16	4.18	1.42	2.54	0.11	0.37	Yes	4.17
4	59.35	49.74	0	0.47	0.27	0.11	8.11	1.42	2.54	0.16	0.28	Yes	5.81
5	85.38	58.38	0	0.53	0.44	0.13	27.58	1.89	8.11	0.18	-0.01	No	9.91

Household taxable income y and consumption c expressed in thousands of USD (2006 values).

6 Summary and conclusion

In this paper, we have evaluated the desirability of child care subsidies in a model where the quantity and quality of care that children receive, both at home (in terms of informal care from parents) and outside the home (at child care facilities), affects the children's human capital that enters into the utility function maximized by parents. Compared to previous contributions in the optimal tax literature, the main distinguishing contribution of our paper has been to emphasize the key importance of the quality dimension of child care (and of the human capital formation process more generally) in determining whether or not child care subsidies are useful as an instrument to achieve redistribution at lower efficiency costs.

We have assessed the case for subsidizing child care by first theoretically analyzing the welfare effects of child care subsidies under different assumptions about the economic variables that are publicly observable at the individual level, and about the policy instruments that are available to the government. As shown in the first part of the paper, these assumptions can play an important role in determining whether child care expenditures should be encouraged or discouraged by the tax system.

We have also assessed the desirability of child care subsidies by means of a calibrated model, incorporating 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 data from the Current Population Survey (CPS) and the American Time Use Survey (ATUS).

In contrast to what has been obtained in previous optimal taxation studies, our theoretical results highlight that it is by no means ex-ante obvious that child care should be subsidized. Moreover, we have provided an empirical calibration where there is no scope for using child care subsidies to reduce the inefficiencies associated with income redistribution, at least when the subsidies are based, as is the case with tax credits, on the child care expenditures incurred by households (rather than on the number of child care hours, as for instance under a public provision scheme). The main reason for our results is that parents choose both the quantity and the quality of center-based child care services, whereas previous studies in the optimal tax literature have analyzed the role of child care subsidies in models where the quality of center-based child care was exogenously given. 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 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.

Subsidies delivered through an opting-out public provision scheme remain, however, 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).

We have focused on the role played by child care subsidies as an instrument for the government to achieve the desired redistributive goals at lower efficiency costs. Another potential argument for subsidizing child care is that there could be externalities associated with the choices of child care arrangements that parents make for their offspring. In Bastani et al. (2017), we considered such aspects by including an externality term in the social welfare function.[50] In such a framework, child care subsidies should be positive provided the externality term carries a sufficiently large weight in the social welfare function. 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.

To conclude, we would like to mention a few potentially broader implications of our results. First, the case for subsidizing child care could be greater in economies that offer a narrower or more homogeneous selection of child care services in the private market, thereby creating a stronger link between child care expenditures and labor supply. Second, one of the most frequently occurring results in applied tax policy discussion is the recommendation that goods complementary to labor should be subsidized, or taxed at lower rates than goods that are complements with leisure, to mitigate the inefficiencies associated with income taxation. Our analysis highlights that such results need to be qualified to take into account the quality dimension of the goods in question.

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.

[50]The externality term can be motivated on the basis that children's 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.

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 L. Micheletto (2017). Child Care Subsidies, Quality, and Optimal Income Taxation. CESifo Working Paper No. 6533, June 2017

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/

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.

Blau, D. and J. Currie (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 Mocan, H. N. (2002). The Supply of Quality in Child Care Centers. *Review of Economics and Statistics*, 84(3), 483-496.

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., 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.

Brewer, M., Saez, E., & Shephard, A. (2010). Means-testing and tax rates on earnings. Dimensions of Tax Design: the mirrlees review, 1.

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.

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.

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.

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.

Herbst, C. M. (2015). The Rising Cost of Child Care in the United States: A Reassessment of the Evidence. IZA Discussion Paper No. 9072.

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

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.

Laughlin, L. (2013). Who's Minding the Kids? Child Care Arrangements: Spring 2011. Current Population Reports, U.S. Census Bureau, Washington, DC.

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.

OECD (2006). Starting Strong II, Paris.

OECD (2011). Doing Better for Families.

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.

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

US Department of Health and Human Services (2009). Child Care Eligibility and Enrolment Estimates for Fiscal Year 2009. Report.

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

Whitehurst, G.J. (2018). What is the market price of daycare and preschool? Evidence Speaks Reports, Vol 2, No. 48, The Brookings Institution.

A Proofs and derivations (online appendix)

A.1 Proof of Proposition 1

Denote by δ the Lagrange multiplier attached to the constraint prescribing a minimum utility level for the high-skilled households, by λ the Lagrange multiplier attached to the self-selection constraint and by μ the Lagrange multiplier attached to the resource constraint of the economy. The first order conditions of the government's program with respect to y^2, c^2, h^2 and q_c^2 are, respectively:

$$(\delta+\lambda)\frac{v'\left(\Theta-\frac{y^2}{w^2}-h^2\right)}{w^2} = \mu(1-\pi) \tag{A1}$$

$$(\delta+\lambda)\,u'\left(c^2\right) = \mu(1-\pi) \tag{A2}$$

$$
\begin{aligned}
&(\delta+\lambda)\left\{\gamma^2\left[\omega^2 f_1'\left(\omega^2 h^2,\left(\Theta-h^2\right)q_c^2\right)-q_c^2 f_2'\left(\omega^2 h^2,\left(\Theta-h^2\right)q_c^2\right)\right]-v'\left(\Theta-\frac{y^2}{w^2}-h^2\right)\right\} \\
= \;& -\mu\left(1-\pi\right)p\left(q_c^2\right) \qquad \text{(A3)}
\end{aligned}
$$

$$
(\delta+\lambda)\gamma^2 f_2'\left(\omega^2 h^2,\left(\Theta-h^2\right)q_c^2\right)=\mu\left(1-\pi\right)p'\left(q_c^2\right) \qquad \text{(A4)}
$$

Using (A2) and taking into account that $\ell^2=\Theta-\frac{y^2}{w^2}-h^2$, one can rewrite conditions (A1), (A3) and (A4) as respectively:

$$
1-\frac{v'\left(\ell^2\right)}{w^2 u'\left(c^2\right)}=0, \qquad \text{(A5)}
$$

$$
p\left(q_c^2\right)+\frac{\gamma^2\left[\omega^2 f_1'\left(\omega^2 h^2,\left(\Theta-h^2\right)q_c^2\right)-q_c^2 f_2'\left(\omega^2 h^2,\left(\Theta-h^2\right)q_c^2\right)\right]}{u'\left(c^2\right)}-\frac{v'\left(\ell^2\right)}{u'\left(c^2\right)}=0, \qquad \text{(A6)}
$$

$$
\frac{\gamma^2 f_2'\left(\omega^2 h^2,\left(\Theta-h^2\right)q_c^2\right)}{u'\left(c^2\right)}-p'\left(q_c^2\right)=0, \qquad \text{(A7)}
$$

i.e. the same kind of conditions that characterize the optimal choices of a high-skilled household under laissez-faire. This shows that no distortion should be imposed, at the solution to the government's program, on the choices of high-skilled households.

The first order conditions with respect to y^1, c^1, h^1 and q_c^1 are instead respectively given by:

$$
\begin{aligned}
\frac{v'\left(\Theta-\frac{y^1}{w^1}-h^1\right)}{w^1} &= \lambda\frac{v'\left(\Theta-\frac{y^1}{w^2}-h^1\right)}{w^2}+\mu\pi \qquad \text{(A8)}\\
u'\left(c^1\right) &= \lambda u'\left(c^1\right)+\mu\pi \qquad \text{(A9)}
\end{aligned}
$$

$$
\begin{aligned}
&\gamma^1\left[\omega^1 f_1'\left(\omega^1 h^1,\left(\Theta-h^1\right)q_c^1\right)-q_c^1 f_2'\left(\omega^1 h^1,\left(\Theta-h^1\right)q_c^1\right)\right]-v'\left(\Theta-\frac{y^1}{w^1}-h^1\right) \\
= \;& \lambda\left\{\gamma^2\left[\omega^2 f_1'\left(\omega^2 h^1,\left(\Theta-h^1\right)q_c^1\right)-q_c^1 f_2'\left(\omega^2 h^1,\left(\Theta-h^1\right)q_c^1\right)\right]-v'\left(\Theta-\frac{y^1}{w^2}-h^1\right)\right\} \\
& -\mu\pi p\left(q_c^1\right) \qquad \text{(A10)}
\end{aligned}
$$

$$
\gamma^1 f_2'\left(\omega^1 h^1,\left(\Theta-h^1\right)q_c^1\right)=\lambda\gamma^2 f_2'\left(\omega^2 h^1,\left(\Theta-h^1\right)q_c^1\right)+\mu\pi p'\left(q_c^1\right) \qquad \text{(A11)}
$$

Combining (A8) and (A9) gives

$$
1-\frac{v'\left(\Theta-\frac{y^1}{w^1}-h^1\right)}{w^1 u'\left(c^1\right)}=\frac{\lambda}{\mu\pi}\left[\frac{v'\left(\Theta-\frac{y^1}{w^1}-h^1\right)}{w^1}-\frac{v'\left(\Theta-\frac{y^1}{w^2}-h^1\right)}{w^2}\right].
$$

Since $w^2>w^1$ implies $\Theta-\frac{y^1}{w^1}-h^1\leq\Theta-\frac{y^1}{w^2}-h^1$ (with $\Theta-\frac{y^1}{w^1}-h^1=\Theta-\frac{y^1}{w^2}-h^1$ only if $y^1=0$), the assumed concavity of the function $v\left(\ell\right)$ ensures that $\frac{v'\left(\Theta-\frac{y^1}{w^1}-h^1\right)}{w^1}-$

$\frac{v'\left(\Theta-\frac{y^1}{w^2}-h^1\right)}{w^2}>0$. Therefore, we can conclude that

$$1-\frac{v'\left(\Theta-\frac{y^1}{w^1}-h^1\right)}{w^1u'\left(c^1\right)}>0.$$

Combining (A10) and (A9) one gets:

$$\begin{aligned}
&p\left(q_c^1\right)+\frac{\gamma^1\left[\omega^1 f_1'\left(\omega^1 h^1,\left(\Theta-h^1\right)q_c^1\right)-q_c^1 f_2'\left(\omega^1 h^1,\left(\Theta-h^1\right)q_c^1\right)\right]}{u'\left(c^1\right)}-\frac{v'\left(\Theta-\frac{y^1}{w^1}-h^1\right)}{u'\left(c^1\right)}\\
=\;&\frac{\lambda}{\mu\pi}\left\{\gamma^2\left[\omega^2 f_1'\left(\omega^2 h^1,\left(\Theta-h^1\right)q_c^1\right)-q_c^1 f_2'\left(\omega^2 h^1,\left(\Theta-h^1\right)q_c^1\right)\right]-v'\left(\Theta-\frac{y^1}{w^2}-h^1\right)\right\}\\
&-\frac{\lambda}{\mu\pi}\left\{\gamma^1\left[\omega^1 f_1'\left(\omega^1 h^1,\left(\Theta-h^1\right)q_c^1\right)-q_c^1 f_2'\left(\omega^1 h^1,\left(\Theta-h^1\right)q_c^1\right)\right]-v'\left(\Theta-\frac{y^1}{w^1}-h^1\right)\right\},
\end{aligned}$$

which implies that it is optimal to distort h^1 downwards (which is equivalent to say that it is optimal to distort h_c^1 upwards) when

$$\begin{aligned}
&\gamma^2\left[\omega^2 f_1'\left(\omega^2 h^1,\left(\Theta-h^1\right)q_c^1\right)-q_c^1 f_2'\left(\omega^2 h^1,\left(\Theta-h^1\right)q_c^1\right)\right]-v'\left(\Theta-\frac{y^1}{w^2}-h^1\right)\\
>\;&\\
&\gamma^1\left[\omega^1 f_1'\left(\omega^1 h^1,\left(\Theta-h^1\right)q_c^1\right)-q_c^1 f_2'\left(\omega^1 h^1,\left(\Theta-h^1\right)q_c^1\right)\right]-v'\left(\Theta-\frac{y^1}{w^1}-h^1\right).
\end{aligned}\tag{A12}$$

When $\gamma^2=\gamma^1$ and $\omega^2=\omega^1$, the condition boils down to

$$v'\left(\Theta-\frac{y^1}{w^1}-h^1\right)-v'\left(\Theta-\frac{y^1}{w^2}-h^1\right)>0,$$

which is indeed always satisfied as long as $y^1>0$.

However, when either $\gamma^2>\gamma^1$ or $\omega^2>\omega^1$ (or both $\gamma^2>\gamma^1$ and $\omega^2>\omega^1$), inequality (A12) might be violated, implying that one cannot rule out the possibility that it is optimal to distort h^1 upwards (which is equivalent to say that it is optimal to distort h_c^1 downwards).[51]

---

[51] Totally differentiating

$$\gamma\left[\omega f_1'\left(\omega h^1,\left(\Theta-h^1\right)q_c^1\right)-q_c^1 f_2'\left(\omega h^1,\left(\Theta-h^1\right)q_c^1\right)\right]-v'\left(\Theta-\frac{y^1}{w}-h^1\right),$$

with respect to γ, ω and w gives:

$$\begin{aligned}
&\left[\omega f_1'\left(\omega h^1,\left(\Theta-h^1\right)q_c^1\right)-q_c^1 f_2'\left(\omega h^1,\left(\Theta-h^1\right)q_c^1\right)\right]d\gamma\\
&+\gamma\left[f_1'\left(\omega h^1,\left(\Theta-h^1\right)q_c^1\right)+\omega h^1 f_{11}''\left(\omega h^1,\left(\Theta-h^1\right)q_c^1\right)-h^1 q_c^1 f_{12}''\right]d\omega\\
&-\frac{y^1}{\left(w\right)^2}v''\left(\Theta-\frac{y^1}{w}-h^1\right)dw,
\end{aligned}$$

Finally, combining (A11) and (A9) one gets:

$$\frac{\gamma^1 f_2'\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right)}{u'\left(b^1\right)} - p'\left(q_c^1\right) = \frac{\lambda}{\mu\pi}\left[\gamma^2 f_2'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right) - \gamma^1 f_2'\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right)\right].$$

When $\gamma^2 = \gamma^1$ and $\omega^2 = \omega^1$ the right hand side of the equation above goes to zero, implying that no distortion should be imposed on q_c^1. Otherwise, if either $\gamma^2 > \gamma^1$ or $\omega^2 > \omega^1$ (or both $\gamma^2 > \gamma^1$ and $\omega^2 > \omega^1$), the right hand side of the equation above is strictly positive, implying that a downward distortion should be imposed on q_c^1.

A.2 Proof of Proposition 2

Denote by δ the Lagrange multiplier attached to the constraint prescribing a minimum utility level for the high-skilled households, by λ the Lagrange multiplier attached to the self-selection constraint and by μ the Lagrange multiplier attached to the resource constraint of the economy. The first order conditions of the government's program with respect to y^2, c^2 and D^2 are, respectively:

$$\left(\delta + \lambda\right)\frac{v'\left(\frac{D^2}{p\left(q_c^2\right)} - \frac{y^2}{w^2}\right)}{w^2} = \mu\left(1 - \pi\right) \tag{A13}$$

$$\left(\delta + \lambda\right) u'\left(c^2\right) = \mu\left(1 - \pi\right) \tag{A14}$$

$$\begin{aligned} & \frac{\delta + \lambda}{p\left(q_c^2\right)}\left[-\omega^2 f_1'\left(\left(\Theta - \frac{D^2}{p\left(q_c^2\right)}\right)\omega^2, \frac{D^2 q_c^2}{p\left(q_c^2\right)}\right) + q_c^2 f_2'\left(\left(\Theta - \frac{D^2}{p\left(q_c^2\right)}\right)\omega^2, \frac{D q_c^2}{p\left(q_c^2\right)}\right)\right]\gamma^2 \\ & + \frac{\delta + \lambda}{p\left(q_c^2\right)} v'\left(\frac{D^2}{p\left(q_c^2\right)} - \frac{y^2}{w^2}\right) \\ = \; & \mu\left(1 - \pi\right) \end{aligned} \tag{A15}$$

---

or equivalently, defining the elasticity $\epsilon_{f_1',h}$ as $\epsilon_{f_1',h} = \frac{df_1'}{dh}\frac{h}{f_1'} = \left(\omega f_{11}'' - q_c f_{12}''\right)\frac{h}{f_1'}$,

$$\begin{aligned} & \left[\omega f_1'\left(\omega h^1, \left(\Theta - h^1\right) q_c^1\right) - q_c^1 f_2'\left(\omega h^1, \left(\Theta - h^1\right) q_c^1\right)\right] d\gamma \\ & + \gamma\left(1 + \epsilon_{f_1',h}\right) f_1'\left(\omega h^1, \left(\Theta - h^1\right) q_c^1\right) d\omega - \frac{y^1}{\left(w\right)^2} v''\left(\Theta - \frac{y^1}{w} - h^1\right) dw, \end{aligned}$$

which is an expression that in general cannot be unambiguously signed.

Using (A14) and taking into account that $\ell^2 = \frac{D^2}{p(q_c^2)} - \frac{y^2}{w^2}$, one can rewrite conditions (A13) and (A15) as, respectively:

$$1 - \frac{v'\left(\ell^2\right)}{w^2 u'\left(c^2\right)} = 0, \tag{A16}$$

$$p\left(q_c^2\right) + \frac{\left[\omega^2 f_1'\left(\left(\Theta - \frac{D^2}{p(q_c^2)}\right)\omega^2, \frac{D^2 q_c^2}{p(q_c^2)}\right) - q_c^2 f_2'\left(\left(\Theta - \frac{D^2}{p(q_c^2)}\right)\omega^2, \frac{D q_c^2}{p(q_c^2)}\right)\right]\gamma^2}{u'\left(c^2\right)} - \frac{v'\left(\ell^2\right)}{u'\left(c^2\right)} = 0. \tag{A17}$$

Taking into account that, for any given value of y^2 and D^2, a high-skilled household chooses q_c^2 to satisfy the condition[52]

$$\left[\omega^2 f_1'\left(\left(\Theta - \frac{D^2}{p\left(q_c^2\right)}\right)\omega^2, \frac{D^2 q_c^2}{p\left(q_c^2\right)}\right) + \left(\frac{p\left(q_c^2\right)}{p'\left(q_c^2\right)} - q_c^2\right) f_2'\left(\left(\Theta - \frac{D^2}{p\left(q_c^2\right)}\right)\omega^2, \frac{D^2 q_c^2}{p\left(q_c^2\right)}\right)\right]\gamma^2 = v'\left(\ell^2\right), \tag{A18}$$

by combining (A17) and (A18) one gets

$$\gamma^2 q_c^2 f_2'\left(\left(\Theta - \frac{D^2}{p\left(q_c^2\right)}\right)\omega^2, \frac{D^2 q_c^2}{p\left(q_c^2\right)}\right) = p'\left(q_c^2\right) u'\left(c^2\right). \tag{A19}$$

Comparing (A16), (A17) and (A19) with the conditions characterizing the optimal behavior of a high-skilled household under laissez-faire, one can see that at the solution to the government's problem all choices by high-skilled households are left undistorted.

The first order conditions of the government's program with respect to y^1, c^1 and D^1 are, respectively:

$$\frac{v'\left(\frac{D^1}{p(q_c^1)} - \frac{y^1}{w^1}\right)}{w^1} = \lambda \frac{v'\left(\frac{D^1}{p\left(\widehat{q_c^1}\right)} - \frac{y^1}{w^2}\right)}{w^2} + \mu\pi \tag{A20}$$

$$u'\left(c^1\right) = \lambda u'\left(c^1\right) + \mu\pi \tag{A21}$$

[52]For given values of y^2 and D^2, a high-skilled household solves the following optimization problem:

$$\max_{q_c^2} \quad u(c) + \gamma^2 f\left(\left(\Theta - \frac{D^2}{p\left(q_c^2\right)}\right)\omega^2, \frac{D^2 q_c^2}{p\left(q_c^2\right)}\right) + v\left(\frac{D^2}{p\left(q_c^2\right)} - \frac{y^2}{w^2}\right).$$

The associated first order condition is

$$\gamma^2\left[\frac{D^2 p'\left(q_c^2\right)}{\left(p\left(q_c^2\right)\right)^2}\omega^2 f_1'\left(\left(\Theta - \frac{D^2}{p\left(q_c^2\right)}\right)\omega^2, \frac{D^2 q_c^2}{p\left(q_c^2\right)}\right) + \left(\frac{D^2}{p\left(q_c^2\right)} - q_c^2\frac{D^2 p'\left(q_c^2\right)}{\left(p\left(q_c^2\right)\right)^2}\right) f_2\right] = \frac{D^2 p'\left(q_c^2\right)}{\left(p\left(q_c^2\right)\right)^2} v'\left(\frac{D^2}{p\left(q_c^2\right)} - \frac{y^2}{w^2}\right).$$

$$\left[-\omega^1 f_1'\left(\left(\Theta-\frac{D^1}{p\left(q_c^1\right)}\right)\omega^1,\frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)+q_c^1 f_2'\left(\left(\Theta-\frac{D^1}{p\left(q_c^1\right)}\right)\omega^1,\frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)\right]\frac{\gamma^1}{p\left(q_c^1\right)}$$
$$+\frac{v'\left(\frac{D^1}{p(q_c^1)}-\frac{y^1}{w^1}\right)}{p\left(q_c^1\right)}$$
$$=\frac{\lambda\gamma^2}{p\left(\widehat{q}_c\right)}\left[-\omega^2 f_1'\left(\left(\Theta-\frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2,\frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)+\widehat{q}_c f_2'\left(\left(\Theta-\frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2,\frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)\right]$$
$$+\frac{\lambda}{p\left(\widehat{q}_c\right)}v'\left(\frac{D^1}{p\left(\widehat{q}_c^2\right)}-\frac{y^1}{w^2}\right)+\mu\pi \tag{A22}$$

Combining (A20) and (A21) gives:

$$1-\frac{v'\left(\frac{D^1}{p(q_c^1)}-\frac{y^1}{w^1}\right)}{w^1 u'\left(c^1\right)}=\frac{\lambda}{\mu\pi}\left[\frac{v'\left(\frac{D^1}{p(q_c^1)}-\frac{y^1}{w^1}\right)}{w^1}-\frac{v'\left(\frac{D^1}{p(\widehat{q}_c)}-\frac{y^1}{w^2}\right)}{w^2}\right]. \tag{A23}$$

Combining (A22) and (A21) gives

$$-p\left(q_c^1\right)+\frac{\left(-\omega^1 f_1+q_c^1 f_2\right)\gamma^1+v'\left(\frac{D^1}{p(q_c^1)}-\frac{y^1}{w^1}\right)}{u'\left(c^1\right)}$$
$$=\frac{\lambda}{\mu\pi}\frac{p\left(q_c^1\right)}{p\left(\widehat{q}_c\right)}\left[\left(-\omega^2 f_1'\left(\left(\Theta-\frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2,\frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)+\widehat{q}_c f_2'\left(\left(\Theta-\frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2,\frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)\right)\gamma^2\right]$$
$$-\frac{\lambda}{\mu\pi}\left[\left(-\omega^1 f_1'\left(\left(\Theta-\frac{D^1}{p\left(q_c^1\right)}\right)\omega^1,\frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)+q_c^1 f_2'\left(\left(\Theta-\frac{D^1}{p\left(q_c^1\right)}\right)\omega^1,\frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)\right)\gamma^1\right]$$
$$+\frac{\lambda p\left(q_c^1\right)}{\mu\pi}\left[\frac{v'\left(\frac{D^1}{p(\widehat{q}_c)}-\frac{y^1}{w^2}\right)}{p\left(\widehat{q}_c\right)}-\frac{v'\left(\frac{D^1}{p(q_c^1)}-\frac{y^1}{w^1}\right)}{p\left(q_c^1\right)}\right]$$

implying that there should be a downward distortion on D^1 iff

$$\left[\left(-\omega^2 f_1'\left(\left(\Theta-\frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2,\frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)+\widehat{q}_c f_2'\left(\left(\Theta-\frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2,\frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)\right)\gamma^2\right]\frac{1}{p\left(\widehat{q}_c\right)}$$
$$+\frac{v'\left(\frac{D^1}{p(\widehat{q}_c)}-\frac{y^1}{w^2}\right)}{p\left(\widehat{q}_c\right)}$$
$$>$$
$$\left[\left(-\omega^1 f_1'\left(\left(\Theta-\frac{D^1}{p\left(q_c^1\right)}\right)\omega^1,\frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)+q_c^1 f_2'\left(\left(\Theta-\frac{D^1}{p\left(q_c^1\right)}\right)\omega^1,\frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)\right)\gamma^1\right]\frac{1}{p\left(q_c^1\right)}$$
$$+\frac{v'\left(\frac{D^1}{p(q_c^1)}-\frac{y^1}{w^1}\right)}{p\left(q_c^1\right)}. \tag{A24}$$

Consider now the first order conditions characterizing an optimal choice for q_c by a low-skilled and a high-skilled household when both choose the bundle intended for low-skilled households. q_c^1 and $\widehat{q}_c$ are chosen to satisfy, respectively:

$$\begin{aligned}\frac{p\left(q_c^1\right)}{p'\left(q_c^1\right)}\gamma^1 f_2'\left(\left(\Theta-\frac{D^1}{p\left(q_c^1\right)}\right)\omega^1,\frac{D^1 q_c^1}{p\left(q_c^1\right)}\right) &= v'\left(\frac{D^1}{p\left(q_c^1\right)}-\frac{y^1}{w^1}\right)\\ &+q_c^1 f_2'\left(\left(\Theta-\frac{D^1}{p\left(q_c^1\right)}\right)\omega^1,\frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)\gamma^1\\ &-\omega^1 f_1'\left(\left(\Theta-\frac{D^1}{p\left(q_c^1\right)}\right)\omega^1,\frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)\gamma^1;\end{aligned} \tag{A25}$$

$$\begin{aligned}\frac{p\left(\widehat{q}_c\right)}{p'\left(\widehat{q}_c\right)}\gamma^2 f_2'\left(\left(\Theta-\frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2,\frac{D^1 \widehat{q}_c}{p\left(\widehat{q}_c\right)}\right) &= v'\left(\frac{D^1}{p\left(\widehat{q}_c\right)}-\frac{y^1}{w^2}\right)\\ &+\widehat{q}_c^2 f_2'\left(\left(\Theta-\frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2,\frac{D^1 \widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)\gamma^2\\ &-\left(\omega^2 f_1'\left(\left(\Theta-\frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2,\frac{D^1 \widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)\right)\gamma^2.\end{aligned} \tag{A26}$$

Substituting (A25)-(A26) into (A24) and simplifying terms, one obtains that there should be a downward distortion on D^1 iff

$$\frac{\gamma^2 f_2'\left(\left(\Theta-\frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2,\frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)}{p'\left(\widehat{q}_c\right)} > \frac{\gamma^1 f_2'\left(\left(\Theta-\frac{D^1}{p\left(q_c^1\right)}\right)\omega^1,\frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)}{p'\left(q_c^1\right)}. \tag{A27}$$

Consider now how, for given values of y and D, the individual optimal choice of q_c is affected by changes in w, γ and ω. The private first order condition characterizing an optimal choice for q_c (for given values of y and D), is given by

$$\frac{p\left(q_c\right)}{p'\left(q_c\right)}\gamma f_2' = v'\left(\frac{D}{p\left(q_c\right)}-\frac{y}{w}\right)-\left(\omega f_1'-q_c f_2'\right)\gamma.$$

Totally differentiating the condition above gives:

$$\begin{aligned}
&\frac{\left(p'\left(q_c\right)\right)^2 - p\left(q_c\right)p''\left(q_c\right)}{\left(p'\left(q_c\right)\right)^2}\gamma f_2' dq_c + \frac{p\left(q_c\right)}{p'\left(q_c\right)}\gamma\left\{\omega\frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}f_{12}'' + \left(\frac{D}{p\left(q_c\right)} - \frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}q_c\right)f_{22}''\right\}dq_c \\
&+\gamma\left\{\omega\frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}f_{11}'' + \left(\frac{D}{p\left(q_c\right)} - \frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}q_c\right)f_{12}''\right\}\omega dq_c \\
&-\gamma\left\{\omega\frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}f_{12}'' + \left(\frac{D}{p\left(q_c\right)} - \frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}q_c\right)f_{22}''\right\}q_c dq_c \\
&+\frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}v'' dq_c - \gamma f_2' dq_c - \frac{y}{\left(w\right)^2}v'' dw \\
&+\gamma f_1' d\omega + \left(\Theta - \frac{D}{p\left(q_c\right)}\right)\gamma\omega f_{11}'' d\omega + \frac{p\left(q_c\right)}{p'\left(q_c\right)}\left(\Theta - \frac{D}{p\left(q_c\right)}\right)\gamma f_{12}'' d\omega - \left(\Theta - \frac{D}{p\left(q_c\right)}\right)\gamma q_c f_{12}'' d\omega \\
&+\omega f_1' d\gamma + \left(\frac{p\left(q_c\right)}{p'\left(q_c\right)} - q_c\right)f_2' d\gamma \\
= \; & 0.
\end{aligned}$$

Define ϵ_{p,q_c} as $\epsilon_{p,q_c} \equiv \frac{p'(q_c)}{p(q_c)}q_c$ and Λ as

$$\begin{aligned}
\Lambda \equiv \; & \left\{\left[\gamma\left(\omega f_{11}'' - q_c f_{12}''\right)\omega + v''\right]\frac{p'\left(q_c\right)}{p\left(q_c\right)} + \left[\left(\frac{1}{\epsilon_{p,q_c}} - 1\right)q_c f_{22}'' + \omega f_{12}''\right]\left(1 - \epsilon_{p,q_c}\right)\gamma\right\}\frac{D}{p\left(q_c\right)} \\
& + \left[\omega\frac{D}{p\left(q_c\right)}f_{12}'' - \frac{p\left(q_c\right)p''\left(q_c\right)}{\left(p'\left(q_c\right)\right)^2}f_2'\right]\gamma, \qquad \text{(A28)}
\end{aligned}$$

where $\Lambda < 0$ from the second order conditions of an individual optimum. It follows that we have:

$$\left(\frac{dq_c}{dw}\right)_{dD=0,dy=0} = \frac{yv''}{\left(w\right)^2\Lambda} > 0 \qquad \text{(A29)}$$

$$\left(\frac{dq_c}{d\omega}\right)_{dD=0,dy=0} = -\frac{f_1' + \left(\Theta - \frac{D}{p(q_c)}\right)\left[\omega f_{11}'' + \left(\frac{1}{\epsilon_{p,q_c}} - 1\right)q_c f_{12}''\right]}{\Lambda}\gamma \qquad \text{(A30)}$$

$$\left(\frac{dq_c}{d\gamma}\right)_{dD=0,dy=0} = -\frac{\omega f_1' + \left(\frac{1}{\epsilon_{p,q_c}} - 1\right)q_c f_2'}{\Lambda}. \qquad \text{(A31)}$$

Denoting by $\epsilon_{f_{1,\omega}'}$ the elasticity of f_1' with respect to ω (i.e. $\epsilon_{f_{1,\omega}'} \equiv \omega h f_{11}''/f_1' = \left(\Theta - \frac{D}{p(q_c)}\right)\omega f_{11}''/f_1'$), we can equivalently rewrite $\left(dq_c/d\omega\right)_{dD=0,dY=0}$ as

$$\left(\frac{dq_c}{d\omega}\right)_{dD=0,dy=0} = -\frac{\left(1 + \epsilon_{f_{1,\omega}'}\right)f_1' + \left(\Theta - \frac{D}{p(q_c)}\right)\left(\frac{1}{\epsilon_{p,q_c}} - 1\right)q_c f_{12}''}{\Lambda}\gamma. \qquad \text{(A32)}$$

Suppose now that agents only differ in terms of wage rates ($\gamma^1 = \gamma^2$ and $\omega^1 = \omega^2$). Since for given values of y and D, leisure ℓ is given by $\frac{D}{p(q_c)} - \frac{y}{w}$, a high-skilled agent behaving as a mimicker will enjoy a higher amount of leisure if, keeping fixed y and D, it is true that $\frac{d\left(\frac{D}{p(q_c)}-\frac{y}{w}\right)}{dw} = \frac{y}{(w)^2} - \frac{Dp'(q_c)}{(p(q_c))^2}\frac{dq_c}{dw} > 0$. Using (A29) we have:

$$\left(\frac{d\left(\frac{D}{p(q_c)} - \frac{y}{w}\right)}{dw}\right)_{dD=0,dy=0} = \frac{(p(q_c))^2\Lambda - Dp'(q_c)v''}{(p(q_c))^2}\frac{1}{\Lambda}\frac{y}{(w)^2},$$

implying that

$$sign\left\{\left(\frac{d\left(\frac{D}{p(q_c)} - \frac{y}{w}\right)}{dw}\right)_{dD=0,dy=0}\right\} = sign\left\{Dp'(q_c)v'' - (p(q_c))^2\Lambda\right\}.$$

Defining ϵ_{p',q_c} as $\epsilon_{p',q_c} \equiv \frac{p''(q_c)}{p'(q_c)}q_c$, and using the definition of Λ provided by (A28), we have:

$$\begin{aligned} Dp'(q_c)v'' - (p(q_c))^2\Lambda &= -(\omega f''_{11} - q_c f''_{12})p'(q_c)\gamma\omega D \\ &\quad - \left(\frac{1-\epsilon_{p,q_c}}{\epsilon_{p,q_c}}q_c f''_{22} + \omega f''_{12}\right)(1-\epsilon_{p,q_c})p(q_c)\gamma D \\ &\quad - \left[\left(1 - \frac{\epsilon_{p',q_c}}{\epsilon_{p,q_c}}\right)f'_2 + \omega\frac{D}{p(q_c)}f''_{12}\right](p(q_c))^2\gamma. \end{aligned}$$

Since each of the three terms on the right hand side of the expression above are non-negative under our initial assumptions on the functions $f(\cdot,\cdot)$ and $p(\cdot)$,[53] we can conclude that, with $\gamma^1 = \gamma^2$ and $\omega^1 = \omega^2$, $\frac{D^1}{p(\widehat{q}_c)} - \frac{y^1}{w^2} > \frac{D^1}{p(q_c^1)} - \frac{y^1}{w^1}$, and therefore, from the concavity of the $v(\cdot)$ function, the right hand side of (A23) is strictly positive, which in turn implies that y^1 is downward distorted $(1 - v'\left(\frac{D^1}{p(q_c^1)} - \frac{y^1}{w^1}\right)/w^1u'(c^1) > 0)$. However, when either $\omega^2 > \omega^1$ or $\gamma^2 > \gamma^1$ (or both $\omega^2 > \omega^1$ and $\gamma^2 > \gamma^1$), one cannot in general rule out the possibility that the right hand side of (A23) is negative, implying that y^1 ought to be distorted upwards.

Consider now the condition determining whether a downward distortion on D^1 is optimal, i.e. condition (A27). We have that, keeping fixed D,

$$\left(\frac{d\left(\frac{\gamma f'_2}{p'(q_c)}\right)}{dw}\right)_{dD=0} = \left(\frac{\partial\frac{\gamma f'_2}{p'(q_c)}}{\partial q_c}\right)_{dD=0}\left(\frac{dq_c}{dw}\right)_{dD=dy=0} \tag{A33}$$

$$\left(\frac{d\left(\frac{\gamma f'_2}{p'(q_c)}\right)}{d\omega}\right)_{dD=0} = \left(\frac{\partial\frac{\gamma f'_2}{p'(q_c)}}{\partial q_c}\right)_{dD=0}\left(\frac{dq_c}{d\omega}\right)_{dD=dy=0} + \left(\Theta - \frac{D}{p(q_c)}\right)\frac{\gamma f''_{12}}{p'(q_c)} \tag{A34}$$

$$\left(\frac{d\left(\frac{\gamma f'_2}{p'(q_c)}\right)}{d\gamma}\right)_{dD=0} = \left(\frac{\partial\frac{\gamma f'_2}{p'(q_c)}}{\partial q_c}\right)_{dD=0}\left(\frac{dq_c}{d\gamma}\right)_{dD=dy=0} + \frac{f'_2}{p'(q_c)}. \tag{A35}$$

[53]For the $f(\cdot,\cdot)$-function we have assumed $f''_{11} < 0$, $f''_{22} < 0$, $f_{12} \geq 0$; for the $p(\cdot)$-function we have assumed that $p(q_c) = k(q_c)^\sigma$, with $k > 0$ and $\sigma \geq 1$. The fact that $\sigma \geq 1$ implies that $1 - \epsilon_{p,q_c} \leq 0$ and $0 < 1 - \frac{\epsilon_{p',q_c}}{\epsilon_{p,q_c}} \leq 1$.

Moreover,

$$\left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0} = \frac{\left[\frac{p'(q_c)}{p(q_c)}\omega f_{12}'' + \left(1 - \frac{p'(q_c)}{p(q_c)}q_c\right) f_{22}''\right] \frac{Dp'(q_c)}{p(q_c)} - f_2' p''\left(q_c\right)}{\left[p'\left(q_c\right)\right]^2}\gamma, \tag{A36}$$

and therefore:

$$\begin{aligned} sign\left\{\left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0}\right\} &= sign\left\{\frac{p'\left(q_c\right)}{p\left(q_c\right)}\omega f_{12}'' + \left(1 - \frac{p'\left(q_c\right)}{p\left(q_c\right)}q_c\right) f_{22}'' - f_2' p''\left(q_c\right)\frac{p\left(q_c\right)}{Dp'\left(q_c\right)}\right\} \\ &= sign\left\{\frac{p'\left(q_c\right)}{p\left(q_c\right)}\omega f_{12}'' + \left(1 - \frac{p'\left(q_c\right)}{p\left(q_c\right)}q_c\right) f_{22}'' - f_2'\frac{p''\left(q_c\right)}{p'\left(q_c\right)}\frac{1}{h_c}\right\} \\ &= sign\left\{\left(\epsilon_{p,q_c}\omega f_{12}'' - \epsilon_{p',q_c}\frac{f_2'}{h_c}\right)\frac{1}{q_c} + \left(1 - \epsilon_{p,q_c}\right) f_{22}''\right\}. \end{aligned} \tag{A37}$$

A.3 Proof of Corollary 1

Part i) Assume, as in Corollary 1, that $f_{12}'' > 0$, $p'' = 0$ (so that $p\left(q_c\right) = kq_c$, $\epsilon_{p,q_c} = 1$ and $\epsilon_{p',q_c} = 0$), $\gamma^1 = \gamma^2$ and $\omega^1 = \omega^2$. It follows from (A37) that $sign\left\{\left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0}\right\} = sign\left\{\omega f_{12}''\frac{1}{q_c}\right\}$, and therefore $\left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0} > 0$. Moreover, from (A29) we have that $\left(\frac{dq_c}{dw}\right)_{dD=dy=0} > 0$ and from (A31) we have that $\left(\frac{dq_c}{d\gamma}\right)_{dD=dy=0} > 0$. Thus, from (A33) and (A35) we have that $\left(\frac{d\left(\frac{\gamma f_2'}{p'(q_c)}\right)}{dw}\right)_{dD=0} > 0$ and $\left(\frac{d\left(\frac{\gamma f_2'}{p'(q_c)}\right)}{d\gamma}\right)_{dD=0} > 0$. To show that condition (A27) is satisfied, and therefore that D^1 should optimally be downward distorted, it is then sufficient to show that $\left(\frac{d\left(\frac{\gamma f_2'}{p'(q_c)}\right)}{d\omega}\right)_{dD=0} > 0$. With $\epsilon_{p,q_c} = 1$, we have that $\left(\frac{dq_c}{d\omega}\right)_{dD=0,dy=0}$ simplifies to:

$$\left(\frac{dq_c}{d\omega}\right)_{dD=0,dy=0} = -\frac{\left(1 + \epsilon_{f_1',\omega}\right) f_1'}{\Lambda}\gamma, \tag{A38}$$

where Λ, defined in (A28), simplifies to:

$$\Lambda = \left[\left(\omega\right)^2 \gamma f_{11}'' + v''\right] \frac{p'\left(q_c\right) D}{\left(p\left(q_c\right)\right)^2}. \tag{A39}$$

Moreover, with $\epsilon_{p,q_c} = 1$, (A36) simplifies to:

$$\left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0} = \frac{\omega f_{12}'' D}{\left[p\left(q_c\right)\right]^2}\gamma. \tag{A40}$$

Therefore, substituting (A38)-(A40) into (A34) gives:

$$\left(\frac{d\left(\frac{\gamma f_2'}{p'(q_c)}\right)}{d\omega}\right)_{dD=0} = -\frac{\left(1+\epsilon_{f_{1,\omega}'}\right)f_1'}{\left[(\omega)^2\gamma f_{11}''+v''\right]p'(q_c)}(\gamma)^2\omega f_{12}''+\frac{\left(\Theta-\frac{D}{p(q_c)}\right)\gamma f_{12}''}{p'(q_c)}$$
$$= \frac{\left(\Theta-\frac{D}{p(q_c)}\right)v''-f_1'\gamma\omega}{(\omega)^2\gamma f_{11}''+v''}\frac{\gamma f_{12}''}{p'(q_c)}>0.$$

Part ii) Now assume that $f_{12}''=f_{22}''=0$, $p''>0$ (so that $p(q_c)=k(q_c)^{\sigma}$, with $\sigma>1$, $\epsilon_{p,q_c}=\sigma$ and $\epsilon_{p',q_c}=\sigma-1>0$), $\gamma^1=\gamma^2$ and $\omega^1=\omega^2$. It follows from (A37) that $sign\left\{\left(\frac{\partial\frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0}\right\}=sign\left\{(1-\sigma)\frac{f_2'}{h_c}\frac{1}{q_c}\right\}$, and therefore $\left(\frac{\partial\frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0}<0$. Moreover, since from (A29) we have that $\left(\frac{dq_c}{dw}\right)_{dD=dY=0}>0$, we can conclude from (A33) that condition (A27) does not hold, implying that D^1 should optimally be upward distorted.

A.4 Proof of Proposition 3

Denote by δ the Lagrange multiplier attached to the constraint prescribing a minimum utility level for the high-skilled households, by λ the Lagrange multiplier attached to the self-selection constraint and by μ the Lagrange multiplier attached to the resource constraint of the economy.

Defining V_β^1, V_β^2 and $\widehat{V}_\beta$ as

$$V_\beta^1\equiv\partial V^1\left(\beta,b^1,y^1\right)/\partial\beta,\quad V_\beta^2\equiv\partial V^2\left(\beta,b^2,y^2\right)/\partial\beta,\quad \widehat{V}_\beta\equiv\partial V^2\left(\beta,b^1,y^1\right)/\partial\beta,$$

the first order conditions of the government's program with respect to y^2, b^2, y^1, b^1 and β are, respectively, given by:

$$(\delta+\lambda)V_y^2 = -\mu(1-\pi)\left\{1-\beta\left[\left(\Theta-h^2\right)p'\left(q_c^2\right)\frac{\partial q_c^2}{\partial y^2}-p\left(q_c^2\right)\frac{\partial h^2}{\partial y^2}\right]\right\}, \quad (A41)$$

$$(\delta+\lambda)V_b^2 = \mu(1-\pi)\left\{1+\beta\left[\left(\Theta-h^2\right)p'\left(q_c^2\right)\frac{\partial q_c^2}{\partial b^2}-p\left(q_c^2\right)\frac{\partial h^2}{\partial b^2}\right]\right\}, \quad (A42)$$

$$V_y^1 = \lambda\widehat{V}_y-\mu\pi\left\{1-\beta\left[\left(\Theta-h^1\right)p'\left(q_c^1\right)\frac{\partial q_c^1}{\partial y^1}-p\left(q_c^1\right)\frac{\partial h^1}{\partial y^1}\right]\right\}, \quad (A43)$$

$$V_b^1 = \lambda\widehat{V}_b^2+\mu\pi\left\{1+\beta\left[\left(\Theta-h^1\right)p'\left(q_c^1\right)\frac{\partial q_c^1}{\partial b^1}-p\left(q_c^1\right)\frac{\partial h^1}{\partial b^1}\right]\right\}, \quad (A44)$$

$$V_\beta^1+(\delta+\lambda)V_\beta^2-\lambda\widehat{V}_\beta-\mu\left\{\pi\left[D^1+\beta\frac{\partial D^1}{\partial\beta}\right]+(1-\pi)\left[D^2+\beta\frac{\partial D^2}{\partial\beta}\right]\right\}=0. \quad (A45)$$

Combining (A41) and (A42) gives

$$\begin{aligned}
&\frac{V_y^2}{V_b^2}\mu\left(1-\pi\right)\left\{1+\beta\left[\left(\Theta-h^2\right)p'\left(q_c^2\right)\frac{\partial q_c^2}{\partial b^2}-p\left(q_c^2\right)\frac{\partial h^2}{\partial b^2}\right]\right\}\\
=\;&-\mu\left(1-\pi\right)\left\{1-\beta\left[\left(\Theta-h^2\right)p'\left(q_c^2\right)\frac{\partial q_c^2}{\partial y^2}-p\left(q_c^2\right)\frac{\partial h^2}{\partial y^2}\right]\right\},
\end{aligned}$$

or, equivalently, and taking into account that $(\Theta-h^2)\,p'\left(q_c^2\right)\frac{\partial q_c^2}{\partial y^2}-p\left(q_c^2\right)\frac{\partial h^2}{\partial y^2}=\frac{\partial D^2}{\partial y^2}$ and $(\Theta-h^2)\,p'\left(q_c^2\right)\frac{\partial q_c^2}{\partial b^2}-p\left(q_c^2\right)\frac{\partial h^2}{\partial b^2}=\frac{\partial D^2}{\partial b^2}$:

$$1+\frac{V_y^2}{V_b^2}=\left(\frac{\partial D^2}{\partial y^2}-\frac{V_y^2}{V_b^2}\frac{\partial D^2}{\partial b^2}\right)\beta. \tag{A46}$$

Given that $-V_y^2/V_b^2$ represents the marginal rate of substitution between y and b for an agent of type 2, the right hand side of (A46) can be rewritten as $\left(\frac{dD^2}{dY^2}\right)_{dV^2=0}\beta$. Moreover, since from the individual optimization problem $\max\limits_{y} V^2\left(\beta,y-T\left(y\right),y\right)$ one can define the implicit marginal income tax rates faced by a high-skilled household as

$$T'=1+V_y^2/V_b^2=1-\frac{v'\left(\Theta-\frac{y^2}{w^2}-h^2\right)}{w^2u'\left(c^2\right)},$$

eq. (A46) can be restated as

$$T'\left(y^2\right)=\left(\frac{dD^2}{dy^2}\right)_{dV^2=0}\beta. \tag{A47}$$

Combining (A43) and (A44) gives

$$\begin{aligned}
&\frac{V_y^1}{V_b^1}\left\{\lambda\widehat{V}_b+\mu\pi\left[1+\beta\left(\left(\Theta-h^1\right)p'\left(q_c^1\right)\frac{\partial q_c^1}{\partial b^1}-p\left(q_c^1\right)\frac{\partial h^1}{\partial b^1}\right)\right]\right\}\\
=\;&\lambda\widehat{V}_y-\mu\pi\left\{1-\beta\left[\left(\Theta-h^1\right)p'\left(q_c^1\right)\frac{\partial q_c^1}{\partial y^1}-p\left(q_c^1\right)\frac{\partial h^1}{\partial y^1}\right]\right\},
\end{aligned}$$

or, equivalently, and taking into account that $(\Theta-h^1)\,p'\left(q_c^1\right)\frac{\partial q_c^1}{\partial y^1}-p\left(q_c^1\right)\frac{\partial h^1}{\partial y^1}=\frac{\partial D^1}{\partial y^1}$ and $(\Theta-h^1)\,p'\left(q_c^1\right)\frac{\partial q_c^1}{\partial b^1}-p\left(q_c^1\right)\frac{\partial h^1}{\partial b^1}=\frac{\partial D^1}{\partial b^1}$:

$$\begin{aligned}
1+\frac{V_y^1}{V_b^1}&=\frac{\lambda\widehat{V}_b}{\mu\pi}\left(\frac{\widehat{V}_y}{\widehat{V}_b}-\frac{V_y^1}{V_b^1}\right)+\beta\left[\frac{\partial D^1}{\partial y^1}-\frac{V_y^1}{V_b^1}\frac{\partial D^1}{\partial b^1}\right]\\
&=\frac{\lambda\widehat{V}_b}{\mu\pi}\left(\frac{\widehat{V}_y}{\widehat{V}_b}-\frac{V_y^1}{V_b^1}\right)+\beta\left(\frac{dD^1}{dy^1}\right)_{dV^1=0}.
\end{aligned}$$

Moreover, since from the individual optimization problem $\max\limits_{y} V^1\left(\beta,y-T\left(y\right),y\right)$ one can define the implicit marginal income tax rates faced by a low-skilled household as

$$T' = 1 + V_y^1 / V_b^1 = 1 - \frac{v'\left(\Theta - \frac{y^1}{w^1} - h^1\right)}{w^1 u'\left(c^1\right)},$$

eq. (A46) can be restated as

$$T'\left(y^1\right) = \frac{\lambda \widehat{V}_b}{\mu \pi}\left(\frac{\widehat{V}_y}{\widehat{V}_b} - \frac{V_y^1}{V_b^1}\right) + \left(\frac{dD^1}{dy^1}\right)_{dV^1=0} \beta. \tag{A48}$$

From Roy's identity we have that

$$V_\beta^1 = D^1 V_b^1, \quad V_\beta^2 = D^2 V_b^2\ , \quad \widehat{V}_\beta = \widehat{D}\widehat{V}_b.$$

Thus, (A45) can be equivalently restated as

$$D^1 V_b^1 + (\delta + \lambda) D^2 V_b^2 - \lambda \widehat{D}\widehat{V}_b - \mu\left\{\pi\left[D^1 + \beta\frac{\partial D^1}{\partial \beta}\right] + (1-\pi)\left[D^2 + \beta\frac{\partial D^2}{\partial \beta}\right]\right\} = 0. \tag{A49}$$

Multiplying (A42) by D^2 and (A44) by D^1 gives:

$$(\delta + \lambda) D^2 V_b^2 = \mu(1-\pi)\left(1 + \beta\frac{\partial D^2}{\partial b^2}\right) D^2, \tag{A50}$$

$$D^1 V_b^1 = \left(\lambda \widehat{V}_b + \mu\pi\right) D^1 + \mu\pi\beta\frac{\partial D^1}{\partial b^1} D^1. \tag{A51}$$

Substituting for $D^1 V_b^1$ and $(\delta + \lambda) D^2 V_b^2$ in (A49) the values provided respectively by (A50) and (A51) gives

$$\lambda\left(D^1 - \widehat{D}\right)\widehat{V}_b - \mu\beta\left\{\left[\frac{\partial D^1}{\partial \beta} - \frac{\partial D^1}{\partial b^1} D^1\right]\pi + \left[\frac{\partial D^2}{\partial \beta} - \frac{\partial D^2}{\partial b^2} D^2\right](1-\pi)\right\} = 0. \tag{A52}$$

Using a tilde symbol to denote compensated (Hicksian) demands, we have that $\frac{\partial \widetilde{D}^i}{\partial \beta} = \frac{\partial D^i}{\partial \beta} - \frac{\partial D^i}{\partial b^i} D^i$ (for $i = 1, 2$). Therefore, it follows from (A52) that

$$\beta = \frac{\lambda \widehat{V}_b}{\mu} \frac{D^1 - \widehat{D}}{\pi\frac{\partial \widetilde{D}^1}{\partial \beta} + (1-\pi)\frac{\partial \widetilde{D}^2}{\partial \beta}}. \tag{A53}$$

A.5 Proof of Corollary 2

Part i) For a given (y, b)-bundle and a proportional subsidy β, an agent characterized by γ, ω and w, will choose h and q_c such that

$$(1-\beta)\, p(q_c)\, u'(c) + \gamma \left[\omega f_1' - q_c f_2'\right] - v' = 0 \tag{A54}$$

$$-(1-\beta)\, p'(q_c)\, u'(c) + \gamma f_2' = 0 \tag{A55}$$

Totally differentiating the system above gives:

$$\begin{aligned}
& (1-\beta)\, p'(q_c)\, u'(c)\, dq_c + \gamma \left[(\omega)^2 f_{11}'' - 2\omega q_c f_{12}'' + (q_c)^2 f_{22}''\right] dh \\
& +\gamma \left[\omega f_{12}'' - q_c f_{22}''\right](\Theta - h)\, dq_c - \gamma f_2' dq_c + v'' dh - \frac{y}{(w)^2} v'' dw \\
& +\gamma \left[f_1' + \omega h f_{11}'' - h q_c f_{12}''\right] d\omega + \left[\omega f_1' - q_c f_2'\right] d\gamma \\
& + (1-\beta)^2 p(q_c)\, u''(c) \left[p(q_c)\, dh - (\Theta - h)\, p'(q_c)\, dq_c\right] \\
= {} & 0
\end{aligned} \tag{A56}$$

$$\begin{aligned}
& \left[-(1-\beta)\, p''(q_c)\, u'(c) + (\Theta - h)\, \gamma f_{22}\right] dq_c + \gamma\left(\omega f_{12} - q_c f_{22}\right) dh \\
& +\gamma h f_{12} d\omega + f_2 d\gamma - (1-\beta)^2 p'(q_c)\, u''(c) \left[p(q_c)\, dh - (\Theta - h)\, p'(q_c)\, dq_c\right] \\
= {} & 0
\end{aligned} \tag{A57}$$

Define Δ_{11}, Δ_{12}, Δ_{21}, Δ_{22} as

$$\begin{aligned}
\Delta_{11} &\equiv \gamma \left[(\omega)^2 f_{11}'' - 2\omega q_c f_{12}'' + (q_c)^2 f_{22}''\right] + v'' + (1-\beta)^2 (p(q_c))^2 u''(c), \\
\Delta_{12} &\equiv \gamma (\Theta - h)\left(\omega f_{12}'' - q_c f_{22}''\right) - (1-\beta)^2 (\Theta - h)\, p'(q_c)\, p(q_c)\, u''(c), \\
\Delta_{21} &\equiv \gamma \left(\omega f_{12}'' - q_c f_{22}''\right) - (1-\beta)^2 p'(q_c)\, p(q_c)\, u''(c), \\
\Delta_{22} &\equiv -(1-\beta)\, p''(q_c)\, u'(c) + (\Theta - h)\, \gamma f_{22}'' + (1-\beta)^2 (\Theta - h)\, (p'(q_c))^2 u''(c).
\end{aligned}$$

Assuming $d\omega = d\gamma = 0$, eqs. (A56)-(A57) can then be expressed in matrix form as

$$\begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix} \begin{bmatrix} dh \\ dq_c \end{bmatrix} = \begin{bmatrix} \frac{y}{(w)^2} v'' dw \\ 0 \end{bmatrix}.$$

Defining by Δ the determinant of the 2X2 matrix above, i.e.

$$\Delta \equiv \Delta_{11}\Delta_{22} - \Delta_{12}\Delta_{21}, \tag{A58}$$

we have that

$$\left(\frac{dh}{dw}\right)_{dy=db=0} = \frac{\left[(1-\beta)^2 (p'(q_c))^2 u''(c) + \gamma f''_{22}\right](\Theta - h)}{\Delta}\frac{yv''}{(w)^2} - \frac{(1-\beta) p''(q_c) u'(c)}{\Delta}\frac{yv''}{(w)^2}, \tag{A59}$$

$$\left(\frac{dq_c}{dw}\right)_{dy=db=0} = \frac{(1-\beta)^2 p'(q_c) p(q_c) u''(c) - (\omega f_{12} - q_c f_{22})\gamma}{\Delta}\frac{yv''}{(w)^2}. \tag{A60}$$

Noticing that $\Delta > 0$ from the second order conditions for an individual optimum, one can then conclude that, based on our assumptions about the functions $p(\cdot)$, $u(\cdot)$, $f(\cdot,\cdot)$ and $v(\cdot)$ (i.e. $p' > 0$, $p'' \geq 0$, $u' > 0$, $u'' < 0$, $f''_{12} \geq 0$, $f''_{22} \leq 0$, $v'' < 0$), $dh/dw > 0$ and $dq_c/dw > 0$.

From (A59)-(A60) we can calculate dD/dw as

$$\begin{aligned}
\left(\frac{dD}{dw}\right)_{dy=db=0} &= -p(q_c)\frac{dh}{dw} + (\Theta - h) p'(q_c)\frac{dq_c}{dw} \\
&= -\frac{p(q_c)\left[-(1-\beta) p''(q_c) u'(c) + (\Theta - h)\gamma f''_{22}\right]}{\Delta}\frac{yv''}{(w)^2} \\
&\quad -\frac{p(q_c)}{\Delta}(1-\beta)^2 (\Theta - h)(p'(q_c))^2 u''(c)\frac{yv''}{(w)^2} \\
&\quad +\frac{(\Theta - h) p'(q_c)}{\Delta}(1-\beta)^2 p'(q_c) p(q_c) u''(c)\frac{yv''}{(w)^2} \\
&\quad -\frac{(\Theta - h) p'(q_c)\gamma(\omega f''_{12} - q_c f''_{22})}{\Delta}\frac{yv''}{(w)^2} \\
&= -\frac{p(q_c)\left[-(1-\beta) p''(q_c) u'(c) + (\Theta - h)\gamma f''_{22}\right]}{\Delta}\frac{yv''}{(w)^2} \\
&\quad -\frac{(\Theta - h) p'(q_c)\gamma(\omega f''_{12} - q_c f''_{22})}{\Delta}\frac{yv''}{(w)^2} \\
&= \frac{(1-\beta) p(q_c) p''(q_c) u'(c) + (\Theta - h)(\epsilon_{p,q_c} - 1) p(q_c)\gamma f''_{22}}{\Delta}\frac{yv''}{(w)^2} \\
&\quad -\frac{(\Theta - h) p'(q_c)\gamma\omega f''_{12}}{\Delta}\frac{yv''}{(w)^2}.
\end{aligned} \tag{A61}$$

With $p'' = 0$, so that $\epsilon_{p,q_c} = 1$, dD/dw simplifies to

$$\left(\frac{dD}{dw}\right)_{dy=db=0} = -\frac{(\Theta - h) p'(q_c)\gamma\omega f''_{12}}{\Delta}\frac{y}{(w)^2}v'' > 0. \tag{A62}$$

Now assume $dw = d\gamma = 0$. From (A56)-(A57) we have

$$\begin{bmatrix}\Delta_{11} & \Delta_{12}\\ \Delta_{21} & \Delta_{22}\end{bmatrix}\begin{bmatrix}dh\\ dq_c\end{bmatrix} = \begin{bmatrix}-(f'_1 + \omega h f''_{11} - h q_c f''_{12})\gamma d\omega\\ -\gamma h f''_{12} d\omega\end{bmatrix},$$

from which one obtains (after some tedious algebra)

$$\begin{aligned}
\left(\frac{dh}{d\omega}\right)_{dy=db=0} &= -\left[\frac{\left(f_1' + \omega h f_{11}'' - h q_c f_{12}''\right) p'\left(q_c\right)}{\Delta} + \frac{h f_{12}''}{\Delta} p\left(q_c\right)\right] (1-\beta)^2 (\Theta - h)\, u''(c)\, \gamma p'\left(q_c\right) \\
&\quad + \frac{\left(f_1' + \omega h f_{11}''\right)\left[(1-\beta) p''\left(q_c\right) u'(c) - (\Theta - h)\gamma f_{22}''\right]}{\Delta}\gamma \\
&\quad + \frac{(\Theta - h) f_{12}'' \gamma\omega - (1-\beta) p''\left(q_c\right) q_c}{\Delta}\gamma h f_{12}'' \qquad \text{(A63)} \\
\left(\frac{dq_c}{d\omega}\right)_{dy=db=0} &= \frac{-v'' h f_{12}'' + \left(\omega f_{12}'' - q_c f_{22}''\right)\gamma f_1' + \left[\left(f_{12}''\right)^2 - f_{11}'' f_{22}''\right]\gamma\omega q_c h}{\Delta}\gamma \\
&\quad - \left[\frac{p\left(q_c\right)}{\Delta} h f_{12}'' + \frac{f_1' + \omega h f_{11}'' - h q_c f_{12}''}{\Delta} p'\left(q_c\right)\right](1-\beta)^2 \gamma p\left(q_c\right) u''(c)\,. \qquad \text{(A64)}
\end{aligned}$$

From (A63)-(A64) we can calculate $dD/d\omega$ as

$$\begin{aligned}
\left(\frac{dD}{d\omega}\right)_{dy=db=0} &= -p\left(q_c\right)\frac{dh}{d\omega} + (\Theta - h) p'\left(q_c\right)\frac{dq_c}{d\omega} \\
&= \frac{h q_c f_{12}'' - \omega h f_{11}'' - f_1'}{\Delta}\gamma p\left(q_c\right)(1-\beta) p''\left(q_c\right) u'(c) \\
&\quad + \frac{(\Theta - h) p\left(q_c\right)\gamma}{\Delta}\left[f_1' f_{22}'' - \omega h \left(f_{12}''\right)^2 + \omega h f_{11}'' f_{22}''\right]\left(1 - \epsilon_{p,q_c}\right)\gamma \\
&\quad + \frac{(\Theta - h) p\left(q_c\right)\gamma}{\Delta}\left(\gamma f_1'\omega - v'' h\right)\frac{\epsilon_{p,q_c}}{q_c} f_{12}''.
\end{aligned}$$

With $p'' = 0$, $dD/d\omega$ simplifies to

$$\left(\frac{dD}{d\omega}\right)_{dy=db=0} = \frac{(\Theta - h)\left(\gamma f_1'\omega - v'' h\right) p\left(q_c\right)\gamma f_{12}''}{q_c \Delta} > 0. \qquad \text{(A65)}$$

Finally, assume $dw = d\omega = 0$. From (A56)-(A57) we have

$$\begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}\begin{bmatrix} dh \\ dq_c \end{bmatrix} = \begin{bmatrix} -\left(\omega f_1' - q_c f_2'\right) d\gamma \\ -f_2' d\gamma \end{bmatrix},$$

from which one obtains (after some tedious algebra)

$$\begin{aligned}
\left(\frac{dh}{d\gamma}\right)_{dy=db=0} &= \frac{\omega f_1'\left[(1-\beta) p''\left(q_c\right) - (\Theta - h)\gamma f_{22}''\right] + f_2'\gamma(\Theta - h)\omega f_{12}'' - q_c f_2'(1-\beta) p''\left(q_c\right)}{\Delta} \\
&\quad - \left[\frac{\left(\omega f_1' - q_c f_2'\right)}{\Delta} p'\left(q_c\right) + \frac{f_2'}{\Delta} p\left(q_c\right)\right](1-\beta)^2 (\Theta - h) u''(c) p'\left(q_c\right), \qquad \text{(A66)} \\
\left(\frac{dq_c}{d\gamma}\right)_{dy=db=0} &= \frac{\left(\omega f_{12}'' - q_c f_{22}''\right)\omega f_1' + \left[\left(f_{12}'' q_c - \omega f_{11}''\right)\omega - \frac{v''}{\gamma}\right] f_2'}{\Delta}\gamma \\
&\quad - \left[\frac{f_2'}{\Delta} p\left(q_c\right) + \frac{\left(\omega f_1' - q_c f_2'\right)}{\Delta} p'\left(q_c\right)\right](1-\beta)^2 u''(c) p\left(q_c\right). \qquad \text{(A67)}
\end{aligned}$$

From (A66)-(A67) we can calculate $dD/d\gamma$ as

$$\begin{aligned}\left(\frac{dD}{d\gamma}\right)_{dy=db=0} &= -p(q_c)\frac{dh}{d\gamma}+(\Theta-h)p'(q_c)\frac{dq_c}{d\gamma} \\ &= \frac{q_c f_2' - \omega f_1'}{\Delta}p(q_c)(1-\beta)p''(q_c)u'(c) \\ &\quad +\frac{(\Theta-h)p(q_c)}{\Delta}(f_1' f_{22}'' - f_2' f_{12}'')(1-\epsilon_{p,q_c})\omega\gamma \\ &\quad +\frac{(\Theta-h)p(q_c)}{\Delta}\left[(\omega)^2(f_1' f_{12}'' - f_2' f_{11}'')\gamma - v'' f_2'\right]\frac{\epsilon_{p,q_c}}{q_c}. \end{aligned} \tag{A68}$$

With $p''=0$, $dD/d\gamma$ simplifies to

$$\left(\frac{dD}{d\gamma}\right)_{dy=db=0} = \frac{(\Theta-h)p(q_c)\left[(\omega)^2(f_1' f_{12}'' - f_2' f_{11}'')\gamma - v'' f_2'\right]}{q_c\Delta} > 0.$$

Based on (A62), (A65) and (A68) we can conclude that, when $f_{12}''>0$, $p''=0$, $\gamma^2 \geq \gamma^1$ and $\omega^2 \geq \omega^1$, we will have that $D^1 < \widehat{D}$. In this case (A53) implies $\beta<0$, i.e. child care expenditures should optimally be taxed rather than subsidized.

Part ii) Assume that $f_{12}'' = f_{22}'' = 0$, $p''>0$, $\gamma^1=\gamma^2$ and $\omega^1=\omega^2$. In this case we have that $sign\left\{\widehat{D}-D^1\right\} = sign\left\{dD/dw\right\}$. Moreover, from (A61) we have that in this case

$$\left(\frac{dD}{dw}\right)_{dy=db=0} = \frac{(1-\beta)p(q_c)p''(q_c)u'(c)}{\Delta}\frac{y}{(w)^2}v'' < 0.$$

Therefore, according to (A53), child care expenditures should optimally be subsidized ($\beta>0$).

A.6 Proof of Corollary 3

Assume that $u''=0$, $p''=0$, $f_{12}''>0$, $\gamma^1=\gamma^2$ and $\omega^1=\omega^2$. From (A62) we already know that $dD/dw>0$, which implies $\beta<0$, i.e. a proportional tax on child care expenditures.

From (A47), and taking into account that $u''=0 \Longrightarrow dD/db=0$, we have

$$T'\left(y^2\right) = \frac{dD^2}{dy^2}\beta.$$

Noticing that

$$\frac{dD}{dy} = -\left(\frac{dD}{dw}\right)_{dy=0}\frac{w}{y}, \tag{A69}$$

from (A62) we have that $p''=0$ implies

$$\frac{dD}{dy} = \frac{(\Theta-h)p'(q_c)\gamma\omega f_{12}''}{\Delta}\frac{1}{w}v'' < 0. \tag{A70}$$

It then follows that $\frac{dD^2}{dy^2}\beta > 0$, which in turn implies $T'(y^2) > 0$.

From (A48), and taking into account that $u'' = 0$ implies $dD/db = 0$ and also $\widehat{V}_b = V_b^1$, we have

$$T'\left(y^1\right) = \frac{\lambda}{\mu\pi}\left(\widehat{V}_y - V_y^1\right) + \frac{dD^1}{dy^1}\beta. \tag{A71}$$

From (A70) and given that $\beta < 0$, we know that $\frac{dD^1}{dy^1}\beta > 0$. Therefore, in order to conclude that $T'(y^1) > 0$, it is sufficient to show that $\widehat{V}_y > V_y^1$, i.e. $-\frac{v'\left(\Theta - \frac{y^1}{w^2} - \widehat{h}\right)}{w^2} > -\frac{v'\left(\Theta - \frac{y^1}{w^1} - h^1\right)}{w^1}$. For this purpose, we will prove that

$$\Theta - \frac{y^1}{w^2} - \widehat{h} > \Theta - \frac{y^1}{w^1} - h^1. \tag{A72}$$

To assess whether the inequality above holds or not, keep fixed y and consider $\left(\frac{d\left(\Theta - \frac{y}{w} - h\right)}{dw}\right)_{dy=0}$. We have:

$$\left(\frac{d\left(\Theta - \frac{y}{w} - h\right)}{dw}\right)_{dy=0} = \frac{y}{(w)^2} - \frac{dh}{dw}.$$

Using (A59) we get (remember that we are now assuming $u'' = 0$ and $p'' = 0$):

$$\left(\frac{d\left(\Theta - \frac{y}{w} - h\right)}{dw}\right)_{dy=0} = \frac{y}{(w)^2}\left[1 - \frac{(\Theta - h)\gamma f_{22}'' v''}{\Delta}\right].$$

It thus follows that $sign\left\{\left(\frac{d\left(\Theta - \frac{y}{w} - h\right)}{dw}\right)_{dy=0}\right\} = sign\left\{1 - \frac{\gamma f_{22}''(\Theta - h)\gamma f_{22}'' v''}{\Delta}\right\}$. Exploiting the definition of Δ provided by (A58), we have (always taking into account that we are assuming $u'' = 0$ and $p'' = 0$):

$$1 - \frac{(\Theta - h)\gamma f_{22}'' v''}{\Delta} = 1 - \frac{\gamma f_{22}'' v''}{\left\{\gamma\left[(\omega)^2 f_{11}'' - 2\omega q_c f_{12}'' + (q_c)^2 f_{22}''\right] + v''\right\}\gamma f_{22}'' - \left[\gamma\left(\omega f_{12}'' - q_c f_{22}''\right)\right]^2},$$

and therefore:

$$1-\frac{\left(\Theta-h\right)\gamma f''_{22}v''}{\Delta} = \frac{\left\{\gamma\left[\left(\omega\right)^{2}f''_{11}-2\omega q_{c}f''_{12}+\left(q_{c}\right)^{2}f''_{22}\right]+v''\right\}\gamma f''_{22}}{\left\{\gamma\left[\left(\omega\right)^{2}f''_{11}-2\omega q_{c}f''_{12}+\left(q_{c}\right)^{2}f''_{22}\right]+v''\right\}\gamma f''_{22}-\left[\gamma\left(\omega f''_{12}-q_{c}f''_{22}\right)\right]^{2}}$$

$$-\frac{\left[\gamma\left(\omega f''_{12}-q_{c}f''_{22}\right)\right]^{2}+\gamma f''_{22}v''}{\left\{\gamma\left[\left(\omega\right)^{2}f''_{11}-2\omega q_{c}f''_{12}+\left(q_{c}\right)^{2}f''_{22}\right]+v''\right\}\gamma f''_{22}-\left[\gamma\left(\omega f''_{12}-q_{c}f''_{22}\right)\right]^{2}}$$

$$= \frac{\left(\omega\right)^{2}f''_{11}f''_{22}-2\omega q_{c}f''_{12}f''_{22}+\left(q_{c}\right)^{2}\left(f''_{22}\right)^{2}}{\left\{\gamma\left[\left(\omega\right)^{2}f''_{11}-2\omega q_{c}f''_{12}+\left(q_{c}\right)^{2}f''_{22}\right]+v''\right\}\gamma f''_{22}-\left[\gamma\left(\omega f''_{12}-q_{c}f''_{22}\right)\right]^{2}}\left(\gamma\right)^{2}$$

$$-\frac{\left(\omega\right)^{2}\left(f''_{12}\right)^{2}+\left(q_{c}\right)^{2}\left(f''_{22}\right)^{2}-2\omega f''_{12}q_{c}f''_{22}}{\left\{\gamma\left[\left(\omega\right)^{2}f''_{11}-2\omega q_{c}f''_{12}+\left(q_{c}\right)^{2}f''_{22}\right]+v''\right\}\gamma f''_{22}-\left[\gamma\left(\omega f''_{12}-q_{c}f''_{22}\right)\right]^{2}}\left(\gamma\right)^{2}$$

$$= \frac{\left[f''_{11}f''_{22}-\left(f''_{12}\right)^{2}\right]\left(\gamma\omega\right)^{2}}{\left\{\gamma\left[\left(\omega\right)^{2}f''_{11}-2\omega q_{c}f''_{12}+\left(q_{c}\right)^{2}f''_{22}\right]+v''\right\}\gamma f''_{22}-\left[\gamma\left(\omega f''_{12}-q_{c}f''_{22}\right)\right]^{2}}.$$

Concavity of the $f\left(\cdot,\cdot\right)$-function $\left(f''_{11}f''_{22}-\left(f''_{12}\right)^{2}>0\right)$ implies $1-\frac{\left(\Theta-h\right)\gamma f''_{22}v''}{\Delta}>0$, and therefore $\left(\frac{d\left(\Theta-\frac{y}{w}-h\right)}{dw}\right)_{dy=0}>0$. In turn, $\left(\frac{d\left(\Theta-\frac{y}{w}-h\right)}{dw}\right)_{dy=0}>0$ implies that (A72) is satisfied. Based on this, we can conclude that the first term appearing on the right hand side of (A71) is positive too, and therefore $T'\left(y^{1}\right)>0$.

A.7 Proof of Corollary 4

The government's problem can then be formally stated as:

$$\max_{y^{1},b^{1},y^{2},b^{2},\beta^{1},\beta^{2}} V^{1}\left(\beta^{1},b^{1},y^{1}\right)$$

subject to

$$V^{2}\left(\beta^{2},b^{2},y^{2}\right)\geq\overline{V},$$

$$V^{2}\left(\beta^{2},b^{2},y^{2}\right)\geq V^{2}\left(\beta^{1},b^{1},y^{1}\right),$$

$$\pi\left(y^{1}-b^{1}-\beta^{1}D^{1}\right)+\left(1-\pi\right)\left(y^{2}-b^{2}-\beta^{2}D^{2}\right)\geq\overline{R},$$

and where $D^{1}\equiv\left(\Theta-h^{1}\right)p\left(q_{c}^{1}\right)$ and $D^{2}\equiv\left(\Theta-h^{2}\right)p\left(q_{c}^{2}\right)$.

Denote by δ the Lagrange multiplier attached to the constraint prescribing a minimum utility level for the high-skilled households, by λ the Lagrange multiplier attached to the self-selection constraint and by μ the Lagrange multiplier attached to the resource constraint of the economy.

The first order conditions of the government's program with respect to y^{2}, b^{2}, β^{2}, y^{1},

b^1, β^1 are, respectively, given by:

$$\begin{aligned}
(\delta+\lambda) V_y^2 &= -\mu(1-\pi)\left\{1-\beta^2\left[\left(\Theta-h^2\right)p'\left(q_c^2\right)\frac{\partial q_c^2}{\partial y^2}-p\left(q_c^2\right)\frac{\partial h^2}{\partial y^2}\right]\right\}, \\
(\delta+\lambda) V_b^2 &= \mu(1-\pi)\left\{1+\beta^2\left[\left(\Theta-h^2\right)p'\left(q_c^2\right)\frac{\partial q_c^2}{\partial b^2}-p\left(q_c^2\right)\frac{\partial h^2}{\partial b^2}\right]\right\},
\end{aligned} \tag{A73}$$

$$(\delta+\lambda) V_{\beta^2}^2-\mu(1-\pi)\left[D^2+\beta^2\frac{\partial D^2}{\partial \beta^2}\right]=0, \tag{A74}$$

$$\begin{aligned}
V_y^1 &= \lambda \widehat{V}_y-\mu\pi\left\{1-\beta^1\left[\left(\Theta-h^1\right)p'\left(q_c^1\right)\frac{\partial q_c^1}{\partial y^1}-p\left(q_c^1\right)\frac{\partial h^1}{\partial y^1}\right]\right\}, \\
V_b^1 &= \lambda \widehat{V}_b^2+\mu\pi\left\{1+\beta^1\left[\left(\Theta-h^1\right)p'\left(q_c^1\right)\frac{\partial q_c^1}{\partial b^1}-p\left(q_c^1\right)\frac{\partial h^1}{\partial b^1}\right]\right\},
\end{aligned} \tag{A75}$$

$$V_{\beta^1}^1-\lambda \widehat{V}_{\beta^1}-\mu\pi\left[D^1+\beta^1\frac{\partial D^1}{\partial \beta}\right]=0. \tag{A76}$$

Applying Roy's identity to (A74) we get:

$$(\delta+\lambda) V_b^2 D^2-\mu(1-\pi)\left[D^2+\beta^2\frac{\partial D^2}{\partial \beta^2}\right]=0,$$

which combined with (A73) gives:

$$\mu(1-\pi)\left\{D^2+\beta^2\left[\left(\Theta-h^2\right)p'\left(q_c^2\right)\frac{\partial q_c^2}{\partial b^2}-p\left(q_c^2\right)\frac{\partial h^2}{\partial b^2}\right]D^2\right\}=\mu(1-\pi)\left[D^2+\beta^2\frac{\partial D^2}{\partial \beta^2}\right],$$

and therefore:

$$\mu(1-\pi)\beta^2\left[\frac{\partial D^2}{\partial \beta^2}-D^2\frac{\partial D^2}{\partial b^2}\right]=0 \Longrightarrow \beta^2=0.$$

Applying Roy's identity to (A76) we get:

$$V_b^1 D^1-\lambda \widehat{V}_b \widehat{D}-\mu\pi\left[D^1+\beta^1\frac{\partial D^1}{\partial \beta}\right]=0,$$

which combined with (A75) gives:

$$\lambda \widehat{V}_b^2 D^1+\mu\pi\left\{D^1+\beta^1\left[\left(\Theta-h^1\right)p'\left(q_c^1\right)\frac{\partial q_c^1}{\partial b^1}-p\left(q_c^1\right)\frac{\partial h^1}{\partial b^1}\right]D^1\right\}=\lambda \widehat{V}_b \widehat{D}+\mu\pi\left[D^1+\beta^1\frac{\partial D^1}{\partial \beta}\right],$$

and therefore:

$$\mu\pi\beta^1\left[\frac{\partial D^1}{\partial \beta}-D^1\frac{\partial D^1}{\partial b^1}\right]=-\lambda \widehat{V}_b\left(\widehat{D}-D^1\right) \Longrightarrow \beta^1=\frac{\lambda \widehat{V}_b}{\mu\pi\frac{\partial \widehat{D}^1}{\partial \beta}}\left(D^1-\widehat{D}\right).$$

A.8 Derivation of (13) and (14)

From condition (11) one obtains:

$$db^1 = -\left(\Theta - h^{1,in}\right) p\left(\overline{q}_c\right) d\beta - \frac{\left[\gamma f_2' - (1-\beta) p' u'\right]\left(\Theta - h^{1,in}\right)}{u'} d\overline{q}_c. \tag{A77}$$

If we now substitute (A77) in (12), this gives:

$$\begin{aligned} & \gamma\left(\omega f_{12}'' - \overline{q}_c f_{22}''\right)\left\{\frac{dh^{1,in}}{d\beta} d\beta + \frac{dh^{1,in}}{d\overline{q}_c} d\overline{q}_c\right\} \\ & -\gamma\left(\omega f_{12}'' - \overline{q}_c f_{22}''\right)\frac{dh^{1,in}}{db^1}\left[\left(\Theta - h^{1,in}\right) p\left(\overline{q}_c\right) d\beta + \frac{\left[\gamma f_2' - (1-\beta) p' u'\right]\left(\Theta - h^{1,in}\right)}{u'} d\overline{q}_c\right] \\ & +\gamma f_{22}\left(\Theta - h^{1,in}\right) d\overline{q}_c \\ = & \; 0, \end{aligned}$$

or, rearranging terms:

$$\begin{aligned} & \gamma\left(\omega f_{12}'' - \overline{q}_c f_{22}''\right)\left[\frac{dh^{1,in}}{d\beta} - \left(\Theta - h^{1,in}\right) p\left(\overline{q}_c\right)\frac{dh^{1,in}}{db^1}\right] d\beta \\ & +\gamma\left(\omega f_{12}'' - \overline{q}_c f_{22}''\right)\left[\frac{dh^{1,in}}{d\overline{q}_c} - \frac{dh^{1,in}}{db^1}\frac{\left[\gamma f_2' - (1-\beta) p' u'\right]\left(\Theta - h^{1,in}\right)}{u'}\right] d\overline{q}_c \\ & +\gamma f_{22}\left(\Theta - h^{1,in}\right) d\overline{q}_c \\ = & \; 0 \end{aligned} \tag{A78}$$

Totally differentiating the first order condition (10), that applies to low-skilled households who opt-in, gives:

$$\frac{dh^{1,in}}{d\beta} = \frac{u'}{v'' + \gamma\left[(\omega)^2 f_{11}'' - 2\omega\overline{q}_c f_{12}'' + \left(\overline{q}_c\right)^2 f_{22}''\right]} p\left(\overline{q}_c\right), \tag{A79}$$

$$\frac{dh^{1,in}}{d\overline{q}_c} = \frac{\left[f_2' - \left(\Theta - h^{1,in}\right)\left(\omega f_{12}'' - \overline{q}_c f_{22}''\right)\right]\gamma - (1-\beta) p' u'}{v'' + \gamma\left[(\omega)^2 f_{11}'' - 2\omega\overline{q}_c f_{12}'' + \left(\overline{q}_c\right)^2 f_{22}''\right]}, \tag{A80}$$

$$\frac{dh^{1,in}}{db^1} = 0. \tag{A81}$$

Substituting (A79)-(A81) in (A78) gives:

$$\begin{aligned} & \gamma\left(\omega f_{12}'' - \overline{q}_c f_{22}''\right)\frac{u'}{v'' + \gamma\left[(\omega)^2 f_{11}'' - 2\omega\overline{q}_c f_{12}'' + \left(\overline{q}_c\right)^2 f_{22}''\right]} p\left(\overline{q}_c\right) d\beta \\ & +\gamma\left(\omega f_{12}'' - \overline{q}_c f_{22}''\right)\frac{\left[f_2' - \left(\Theta - h^{1,in}\right)\left(\omega f_{12}'' - \overline{q}_c f_{22}''\right)\right]\gamma - (1-\beta) p' u'}{v'' + \gamma\left[(\omega)^2 f_{11}'' - 2\omega\overline{q}_c f_{12}'' + \left(\overline{q}_c\right)^2 f_{22}''\right]} d\overline{q}_c \\ & +\gamma f_{22}\left(\Theta - h^{1,in}\right) d\overline{q}_c = 0, \end{aligned}$$

from which one obtains

$$d\overline{q}_c = \frac{\left(\omega f_{12}'' - \overline{q}_c f_{22}''\right) p\left(\overline{q}_c\right) u'}{\left\{\gamma\left(\omega\right)^2\left[\left(f_{12}''\right)^2 - f_{11}'' f_{22}''\right] - v'' f_{22}\right\}\left(\Theta - h^{1,in}\right) - \left(\omega f_{12}'' - \overline{q}_c f_{22}''\right)\left[\gamma f_2' - \left(1-\beta\right) p' u'\right]} d\beta. \tag{A82}$$

Substituting the value found above for $d\overline{q}_c$ into (A77) gives:

$$db^1 = -\left\{\frac{\gamma\left(\omega\right)^2\left[\left(f_{12}''\right)^2 - f_{11}'' f_{22}''\right] - v'' f_{22}''}{\gamma\left(\omega\right)^2\left[\left(f_{12}''\right)^2 - f_{11}'' f_{22}''\right] - v'' f_{22}'' - \frac{\left(\omega f_{12}'' - \overline{q}_c f_{22}''\right)\left[\gamma f_2' - (1-\beta) p'\left(\overline{q}_c\right) u'\right]}{\Theta - h^{1,in}}}\right\}\left(\Theta - h^{1,in}\right) p\left(\overline{q}_c\right) d\beta. \tag{A83}$$

Finally, taking into account that at the initial equilibrium, $\beta = 0$ and $\overline{q} = q_c^{1*}$, so that $\gamma f_2' - (1-\beta) p'\left(\overline{q}_c\right) u' = 0$, eqs. (A82)-(A83) can be simplified to obtain (13)-(14).

A.9 Derivation of the negative welfare effect of an opting-out public provision scheme on the utility of mimicking households

Denote the Lagrange multiplier attached to the first self-selection constraint (i.e. the one prescribing high-skilled not to mimic and opt-out) by λ, and the multiplier attached to the second self-selection constraint (i.e. the one prescribing high-skilled not to mimic and opt-in) by λ^{in}. With respect to the impact on the first self-selection constraint we have a positive mimicking-deterring effect since mimickers' utility change by

$$dV^2\left(b^1, y^1\right) = -p\left(\overline{q}_c\right)\left(\Theta - h^{1,in}\right)\frac{\partial V^2\left(b^1, y^1\right)}{\partial b^1} d\beta < 0.$$

With respect to the impact on the second self-selection constraint we have that mimickers' utility change by

$$\begin{aligned}
& dV^2\left(\overline{q}_c,\beta,b^1,y^1\right) \\
= & \frac{\partial V^2\left(\overline{q}_c,\beta,b^1,y^1\right)}{\partial \overline{q}_c}d\overline{q}_c+\frac{\partial V^2\left(\overline{q}_c,\beta,b^1,y^1\right)}{\partial \beta}d\beta+\frac{\partial V^2\left(\overline{q}_c,\beta,b^1,y^1\right)}{\partial b^1}db^1 \\
= & \left(\Theta-\widehat{h}^{in}\right)\gamma \widehat{f}_2'^{in}\left(\omega\widehat{h}^{in},\left(\Theta-\widehat{h}^{in}\right)\overline{q}_c\right)\frac{\left(\omega f_{12}''-\overline{q}_c f_{22}''\right)p\left(\overline{q}_c\right)u'}{\gamma\left(\Theta-h^{1,in}\right)(\omega)^2\left[\left(f_{12}''\right)^2-f_{11}''f_{22}''\right]-\left(\Theta-h^{1,in}\right)v''f_{22}''}d\beta \\
& -\left(\Theta-\widehat{h}^{in}\right)(1-\beta)p'\left(\overline{q}_c\right)u'\frac{\left(\omega f_{12}''-\overline{q}_c f_{22}''\right)p\left(\overline{q}_c\right)u'}{\gamma\left(\Theta-h^{1,in}\right)(\omega)^2\left[\left(f_{12}''\right)^2-f_{11}''f_{22}''\right]-\left(\Theta-h^{1,in}\right)v''f_{22}''}d\beta \\
& +p\left(\overline{q}_c\right)\left(\Theta-\widehat{h}^{in}\right)u'd\beta-\left(\Theta-h^{1,in}\right)p\left(\overline{q}_c\right)u'd\beta \\
= & \left[\gamma \widehat{f}_2'^{in}\left(\omega\widehat{h}^{in},\left(\Theta-\widehat{h}^{in}\right)\overline{q}_c\right)-(1-\beta)p'\left(\overline{q}_c\right)u'\right]\frac{\Theta-\widehat{h}^{in}}{\Theta-h^{1,in}}\frac{\left(\omega f_{12}''-\overline{q}_c f_{22}''\right)p\left(\overline{q}_c\right)u'}{\gamma(\omega)^2\left[\left(f_{12}''\right)^2-f_{11}''f_{22}''\right]-v''f_{22}''}d\beta \\
& +p\left(\overline{q}_c\right)\left(h^{1,in}-\widehat{h}^{in}\right)u'd\beta. \qquad \text{(A84)}
\end{aligned}$$

To assess the sign of the expression above one needs to determine the sign of $h^{1,in}-\widehat{h}^{in}$. For this purpose, consider the first order condition characterizing the private optimal choice of h for a household who opts-in:

$$\begin{aligned}
& (1-\beta)p\left(\overline{q}_c\right)u'\left(b^1-(1-\beta)(\Theta-h)p\left(\overline{q}_c\right)\right)-v'\left(\Theta-\frac{y^1}{w}-h\right) \\
& +\gamma\left[\omega f_1'\left(\omega h,(\Theta-h)\overline{q}_c\right)-\overline{q}_c f_2'\left(\omega h,(\Theta-h)\overline{q}_c\right)\right] \\
= & \ 0.
\end{aligned}$$

Totally differentiating the first order condition above gives (and taking into account that we are here assuming $u''=0$):

$$v''dh+\gamma\left[(\omega)^2 f_{11}''-2\omega\overline{q}_c f_{12}''+\left(\overline{q}_c\right)^2 f_{22}''\right]dh-\frac{y}{(w)^2}v''dw=0$$

Thus, defining Υ as

$$\Upsilon\equiv v''+\gamma\left[(\omega)^2 f_{11}''-2\omega\overline{q}_c f_{12}''+\left(\overline{q}_c\right)^2 f_{22}''\right]<0,$$

we have:

$$\frac{dh}{dw}=\frac{yv''}{(w)^2\Upsilon}>0,$$

which in turn allows us to conclude that $h^{1,in}-\widehat{h}^{in}<0$.

Thus, the last term on the right hand side of (A84) is negative. Regarding the other term, its sign is the opposite of the sign of the expression within square brackets. However, since we know that at the pre-reform equilibrium $\overline{q}_c$ satisfied $p'\left(\overline{q}_c\right)=\frac{\gamma f_2'\left(\omega h^{1,in},\left(\Theta-h^{1,in}\right)\overline{q}_c\right)}{u'}$, having established that $h^{1,in}-\widehat{h}^{in}<0$ allows concluding that $\gamma \widehat{f}_2'^{in}\left(\omega\widehat{h}^{in},\left(\Theta-\widehat{h}^{in}\right)\overline{q}_c\right)-$

$(1-\beta)\, p'\left(\overline{q}_c\right) u' > 0$. We can then conclude that the proposed reform also has a detrimental effect on a high-skilled households who mimic and opt in.

B Child care subsidies in the United States (online appendix)

Focusing on the case of a family with one child filing jointly, in this appendix we describe in more detail the rules governing the federal and state subsidies that we model in our analysis.

At the federal level there are two tax credits. One (the CTC, i.e. Child Tax Credit) 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 (the CDCTC, i.e. Child and Dependent Care 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 8.

Table 8: 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%		

Since US states usually offers an additional tax credit that differs in generosity across states, in our analysis we set focus on the case of California and model the California child care tax credit which is a fraction of the second federal tax credit illustrated above. (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:

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

where $\beta^{CAL}\left(y^{AGI}\right)$ takes value between 0% and 50% according to the decreasing schedule in table 8.

Finally, the last subsidy scheme that we model is the CCDF (Child Care and Development Fund). This is a block grant fund managed by states within certain federal guidelines. CCDF subsidies are available as vouchers or as part of direct purchase programs, and is primarily targeted to low income families (eligibility is restricted to families with income below 85% of the state median income) who are engaged in work related activities. Whereas the federal recommended subsidy rate for the CCDF is 90%, only a certain proportion of eligible households (those with income below 85% of state median income) receive the subsidy: 52%, 37%, and 18% of households (with kids aged under 6) living, respectively, below, between 101% and 150%, and above 150% of the poverty threshold (US Department of Health and Human Services, 2009). Based on these figures, and considering a baseline CCDF rate equal to 90%, which is the recommended subsidy rate under Federal guidelines, we therefore approximate the CCDF effective subsidy rate (for a family with two adults filing jointly and one kid aged under 6) through a linearly decreasing function that starts at 97% (when the household AGI is equal to zero) and reaches zero when the household AGI is equal to 41.000 USD (where 41.000 USD represents the eligibility threshold in California, defined as 247% of the poverty threshold).

C Computational approach

The optimal tax problem that we solve in this paper is a so-called bi-level programming problem. 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.[54] 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 numeri-

[54] 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 four privately chosen variables that are handled in the subproblem. These are: the labor supply of the mother, the hours of maternal care, the hours of formal child care, and the quality in formal care.

cally 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 implementation of the key computations in 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.[55]

In the main text, we describe the main simplifications that we have adopted. These 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 suffi-

[55]For a discussion about the exponential increase in the number of self-selection constraints in a multi-dimensional 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.

ciently 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 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. 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.[56]

D Robustness with respect to specification of innate ability (online appendix)

In table 9 we show the results for the means-tested subsidy for the case where the innate ability of the child is given by $\gamma^i = \frac{(w_m^i + w_f^i)/2}{\sum_{i=1}^{N}(w_m^i + w_f^i)/2}$.

[56]The model was solved on a dual processor Intel Xeon workstation with a large number of computational cores.

Table 9: Means-tested subsidy

Allocation in households where the mother works

i	y	c	L_m	L_f	h_m	$\frac{D}{y}$	q	q_c	$p(q_c)$	$\frac{T}{y}$	$T'(y)$	β	U
1	47.17	44.63	0.21	0.47	0.09	0.13	1.38	1.26	1.53	-0.08	0.17	-0.06	2.41
2	58.05	50.6	0.27	0.44	0.11	0.13	2.51	1.34	2	-0.01	0.24	-0.08	3.39
3	79.21	62.72	0.35	0.46	0.12	0.13	4.24	1.46	2.87	0.08	0.14	-0.03	4.54
4	95.21	72	0.37	0.42	0.15	0.13	8.16	1.6	4.12	0.11	0.17	0	6.31
5	143.84	103.86	0.45	0.4	0.17	0.13	22.05	1.85	7.47	0.15	-0	0	10.41

Allocation in households where the mother does not work

i	y	c	L_m	L_f	h_m	$\frac{D}{y}$	q	q_c	$p(q_c)$	$\frac{T}{y}$	$T'(y)$	β	U
1	38.62	38.07	0	0.47	0.16	0.13	1.42	1.22	1.36	-0.13	0.24	-0.12	2.28
2	45.55	41.41	0	0.49	0.22	0.13	2.65	1.33	1.96	-0.05	0.22	-0.07	3.14
3	49.21	43.25	0	0.46	0.28	0.14	4.68	1.42	2.54	-0.02	0.31	-0.08	4.21
4	61.56	49.49	0	0.49	0.34	0.14	8.79	1.59	4.01	0.05	0.2	0	5.77
5	83.94	64.39	0	0.52	0.43	0.13	23.69	1.86	7.71	0.1	0	0	9.58

Household taxable income y and consumption c expressed in thousands of USD (2006 values).