Online Appendix for Child Care Subsidies, Quality, and Optimal Income Taxation

Spencer Bastani[*]   Sören Blomquist[†]   Luca Micheletto[‡]

December 8, 2019

A Case 4: Earned income taxed non-linearly, child care expenditures subsidized or taxed at a proportional rate

Assume 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 subsection 3.2, 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. 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.[1]

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

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

[*]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.

[1]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 (problem Þ4):

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

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. $\widehat{D} \equiv (\Theta - \widehat{h}) p(\widehat{q}_c)$) 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 1. *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:*

$$
1 - \frac{v'\left(\Theta - \frac{y^2}{w^2} - h^2\right)}{w^2 u'(c^2)} = T'\left(y^2\right) = \left(\frac{dD^2}{dy^2}\right)_{dV^2=0} \beta. \tag{A1}
$$

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:

$$
1 - \frac{v'\left(\Theta - \frac{y^1}{w^1} - h^1\right)}{w^1 u'(c^1)} = T'\left(y^1\right) = \left(\frac{dD^1}{dy^1}\right)_{dV^1=0} \beta + \frac{\lambda \widehat{V}_b}{\mu \pi}\left(\frac{\widehat{V}_y}{\widehat{V}_b} - \frac{V_y^1}{V_b^1}\right). \tag{A2}
$$

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 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:

$$(\delta + \lambda) \, V_y^2 = -\mu \, (1 - \pi) \left\{ 1 - \beta \left[\left(\Theta - h^2 \right) p' \left(q_c^2 \right) \frac{\partial q_c^2}{\partial y^2} - p \left(q_c^2 \right) \frac{\partial h^2}{\partial y^2} \right] \right\}, \quad \text{(A3)}$$

$$(\delta + \lambda) \, V_b^2 = \mu \, (1 - \pi) \left\{ 1 + \beta \left[\left(\Theta - h^2 \right) p' \left(q_c^2 \right) \frac{\partial q_c^2}{\partial b^2} - p \left(q_c^2 \right) \frac{\partial h^2}{\partial b^2} \right] \right\}, \quad \text{(A4)}$$

$$V_y^1 = \lambda \widehat{V}_y - \mu \pi \left\{ 1 - \beta \left[\left(\Theta - h^1 \right) p' \left(q_c^1 \right) \frac{\partial q_c^1}{\partial y^1} - p \left(q_c^1 \right) \frac{\partial h^1}{\partial y^1} \right] \right\}, \quad \text{(A5)}$$

$$V_b^1 = \lambda \widehat{V}_b^2 + \mu \pi \left\{ 1 + \beta \left[\left(\Theta - h^1 \right) p' \left(q_c^1 \right) \frac{\partial q_c^1}{\partial b^1} - p \left(q_c^1 \right) \frac{\partial h^1}{\partial b^1} \right] \right\}, \quad \text{(A6)}$$

$$V_\beta^1 + (\delta + \lambda) \, V_\beta^2 - \lambda \widehat{V}_\beta - \mu \left\{ \pi \left[D^1 + \beta \frac{\partial D^1}{\partial \beta} \right] + (1 - \pi) \left[D^2 + \beta \frac{\partial D^2}{\partial \beta} \right] \right\} = 0. \quad \text{(A7)}$$

Combining (A3) and (A4) gives

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

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

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

$$1 + \frac{V_y^2}{V_b^2} = \left(\frac{\partial D^2}{\partial y^2} - \frac{V_y^2}{V_b^2} \frac{\partial D^2}{\partial b^2} \right) \beta. \quad \text{(A8)}$$

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 (A8) can be rewritten as $\left(\frac{dD^2}{dY^2} \right)_{dV^2 = 0} \beta$. Moreover,

since from the individual optimization problem $\max_{y} V^2\left(\beta, y - T\left(y\right), y\right)$ one can define the implicit marginal income tax rates faced by a high-skilled household as

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

eq. (A8) can be restated as

$$T'\left(y^2\right) = \left(\frac{dD^2}{dy^2}\right)_{dV^2=0} \beta. \tag{A9}$$

Combining (A5) and (A6) gives

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

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

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

Moreover, since from the individual optimization problem $\max_{y} V^1\left(\beta, y - T\left(y\right), y\right)$ 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. (A8) can be restated as

$$T'\left(y^1\right) = \frac{\lambda\widehat{V_b}}{\mu\pi}\left(\frac{\widehat{V_y}}{\widehat{V_b}} - \frac{V_y^1}{V_b^1}\right) + \left(\frac{dD^1}{dy^1}\right)_{dV^1=0} \beta. \tag{A10}$$

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, (A7) can be equivalently restated as

$$D^1 V_b^1 + \left(\delta + \lambda\right)D^2 V_b^2 - \lambda\widehat{D}\widehat{V_b} - \mu\left\{\pi\left[D^1 + \beta\frac{\partial D^1}{\partial\beta}\right] + \left(1 - \pi\right)\left[D^2 + \beta\frac{\partial D^2}{\partial\beta}\right]\right\} = 0. \tag{A11}$$

Multiplying (A4) by D^2 and (A6) by D^1 gives:

$$(\delta + \lambda) D^2 V_b^2 = \mu (1 - \pi) \left(1 + \beta \frac{\partial D^2}{\partial b^2} \right) D^2, \tag{A12}$$

$$D^1 V_b^1 = \left(\lambda \widehat{V_b} + \mu \pi \right) D^1 + \mu \pi \beta \frac{\partial D^1}{\partial b^1} D^1. \tag{A13}$$

Substituting for $D^1 V_b^1$ and $(\delta + \lambda) D^2 V_b^2$ in (A11) the values provided respectively by (A12) and (A13) gives

$$\lambda \left(D^1 - \widehat{D} \right) \widehat{V_b} - \mu \beta \left\{ \left[\frac{\partial D^1}{\partial \beta} - \frac{\partial D^1}{\partial b^1} D^1 \right] \pi + \left[\frac{\partial D^2}{\partial \beta} - \frac{\partial D^2}{\partial b^2} D^2 \right] (1 - \pi) \right\} = 0. \tag{A14}$$

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 (A14) that

$$\beta = \frac{\lambda \widehat{V_b}}{\mu} \frac{D^1 - \widehat{D}}{\pi \frac{\partial \widetilde{D}^1}{\partial \beta} + (1 - \pi) \frac{\partial \widetilde{D}^2}{\partial \beta}}. \tag{A15}$$

□

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).[2]

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 1. *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 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

$$(1 - \beta) p(q_c) u'(c) + \gamma [\omega f'_1 - q_c f'_2] - v' = 0 \tag{A16}$$
$$-(1 - \beta) p'(q_c) u'(c) + \gamma f'_2 = 0 \tag{A17}$$

Totally differentiating the system above gives:

$$
\begin{aligned}
(1 - \beta) p'(q_c) u'(c) dq_c + \gamma \left[(\omega)^2 f''_{11} - 2\omega q_c f''_{12} + (q_c)^2 f''_{22} \right] dh \\
+ \gamma [\omega f''_{12} - q_c f''_{22}] (\Theta - h) dq_c - \gamma f'_2 dq_c + v'' dh - \frac{y}{(w)^2} v'' dw \\
+ \gamma [f'_1 + \omega h f''_{11} - h q_c f''_{12}] d\omega + [\omega f'_1 - q_c f'_2] d\gamma \\
+ (1 - \beta)^2 p(q_c) u''(c) [p(q_c) dh - (\Theta - h) p'(q_c) dq_c] \\
= 0
\end{aligned}
\tag{A18}
$$

$$
\begin{aligned}
[-(1 - \beta) p''(q_c) u'(c) + (\Theta - h) \gamma f_{22}] dq_c + \gamma (\omega f_{12} - q_c f_{22}) dh \\
+ \gamma h f_{12} d\omega + f_2 d\gamma - (1 - \beta)^2 p'(q_c) u''(c) [p(q_c) dh - (\Theta - h) p'(q_c) dq_c] \\
= 0
\end{aligned}
\tag{A19}
$$

---

[2]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 $\frac{\partial \widetilde{D}^i}{\partial \beta} = \frac{\partial D^i}{\partial \beta}$.

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

$$\Delta_{11} \equiv \gamma \left[(\omega)^2 f_{11}'' - 2\omega q_c f_{12}'' + (q_c)^2 f_{22}'' \right] + v'' + (1-\beta)^2 (p(q_c))^2 u''(c),$$

$$\Delta_{12} \equiv \gamma (\Theta - h) (\omega f_{12}'' - q_c f_{22}'') - (1-\beta)^2 (\Theta - h) p'(q_c) p(q_c) u''(c),$$

$$\Delta_{21} \equiv \gamma (\omega f_{12}'' - q_c f_{22}'') - (1-\beta)^2 p'(q_c) p(q_c) u''(c),$$

$$\Delta_{22} \equiv -(1-\beta) p''(q_c) u'(c) + (\Theta - h) \gamma f_{22}'' + (1-\beta)^2 (\Theta - h) (p'(q_c))^2 u''(c).$$

Assuming $d\omega = d\gamma = 0$, eqs. (A18)-(A19) 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.

$$\Delta \equiv \Delta_{11} \Delta_{22} - \Delta_{12} \Delta_{21}, \tag{A20}$$

we have that

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

$$- \frac{(1-\beta) p''(q_c) u'(c)}{\Delta} \frac{yv''}{(w)^2}, \tag{A21}$$

$$\left(\frac{dq_c}{dw} \right)_{dy=db=0} = \frac{(1-\beta)^2 p'(q_c) p(q_c) u''(c) - (\omega f_{12} - q_c f_{22}) \gamma}{\Delta} \frac{yv''}{(w)^2}. \tag{A22}$$

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 (A21)-(A22) we can calculate dD/dw as

$$
\begin{aligned}
\left(\frac{dD}{dw}\right)_{dy=db=0} &= -p\left(q_c\right)\frac{dh}{dw} + \left(\Theta - h\right) p'\left(q_c\right)\frac{dq_c}{dw} \\
&= -\frac{p\left(q_c\right)\left[-\left(1-\beta\right)p''\left(q_c\right)u'\left(c\right) + \left(\Theta - h\right)\gamma f_{22}''\right]}{\Delta}\frac{yv''}{\left(w\right)^2} \\
&\quad -\frac{p\left(q_c\right)}{\Delta}\left(1-\beta\right)^2\left(\Theta - h\right)\left(p'\left(q_c\right)\right)^2 u''\left(c\right)\frac{yv''}{\left(w\right)^2} \\
&\quad +\frac{\left(\Theta - h\right)p'\left(q_c\right)}{\Delta}\left(1-\beta\right)^2 p'\left(q_c\right)p\left(q_c\right)u''\left(c\right)\frac{yv''}{\left(w\right)^2} \\
&\quad -\frac{\left(\Theta - h\right)p'\left(q_c\right)\gamma\left(\omega f_{12}'' - q_c f_{22}''\right)}{\Delta}\frac{yv''}{\left(w\right)^2} \\
&= -\frac{p\left(q_c\right)\left[-\left(1-\beta\right)p''\left(q_c\right)u'\left(c\right) + \left(\Theta - h\right)\gamma f_{22}''\right]}{\Delta}\frac{yv''}{\left(w\right)^2} \\
&\quad -\frac{\left(\Theta - h\right)p'\left(q_c\right)\gamma\left(\omega f_{12}'' - q_c f_{22}''\right)}{\Delta}\frac{yv''}{\left(w\right)^2} \\
&= \frac{\left(1-\beta\right)p\left(q_c\right)p''\left(q_c\right)u'\left(c\right) + \left(\Theta - h\right)\left(\epsilon_{p,q_c} - 1\right)p\left(q_c\right)\gamma f_{22}''}{\Delta}\frac{yv''}{\left(w\right)^2} \\
&\quad -\frac{\left(\Theta - h\right)p'\left(q_c\right)\gamma\omega f_{12}''}{\Delta}\frac{yv''}{\left(w\right)^2}. \tag{A23}
\end{aligned}
$$

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

$$
\left(\frac{dD}{dw}\right)_{dy=db=0} = -\frac{\left(\Theta - h\right)p'\left(q_c\right)\gamma\omega f_{12}''}{\Delta}\frac{y}{\left(w\right)^2}v'' > 0. \tag{A24}
$$

Now assume $dw = d\gamma = 0$. From (A18)-(A19) we have

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

from which one obtains (after some tedious algebra)

$$
\begin{aligned}
\left(\frac{dh}{d\omega}\right)_{dy=db=0} &= -\left[\frac{\left(f_1' + \omega h f_{11}'' - h q_c f_{12}''\right)p'\left(q_c\right)}{\Delta} + \frac{h f_{12}''}{\Delta}p\left(q_c\right)\right]\left(1-\beta\right)^2\left(\Theta - h\right)u''\left(c\right)\gamma p'\left(q_c\right) \\
&\quad +\frac{\left(f_1' + \omega h f_{11}''\right)\left[\left(1-\beta\right)p''\left(q_c\right)u'\left(c\right) - \left(\Theta - h\right)\gamma f_{22}''\right]}{\Delta}\gamma \\
&\quad +\frac{\left(\Theta - h\right)f_{12}''\gamma\omega - \left(1-\beta\right)p''\left(q_c\right)q_c}{\Delta}\gamma h f_{12}'' \tag{A25} \\
\left(\frac{dq_c}{d\omega}\right)_{dy=db=0} &= \frac{-v'' h f_{12}'' + \left(\omega f_{12}'' - q_c f_{22}''\right)\gamma f_1' + \left[\left(f_{12}''\right)^2 - f_{11}'' f_{22}''\right]\gamma\omega q_c h}{\Delta}\gamma \\
&\quad -\left[\frac{p\left(q_c\right)}{\Delta}h f_{12}'' + \frac{f_1' + \omega h f_{11}'' - h q_c f_{12}''}{\Delta}p'\left(q_c\right)\right]\left(1-\beta\right)^2\gamma p\left(q_c\right)u''\left(c\right). \tag{A26}
\end{aligned}
$$

From (A25)-(A26) we can calculate $dD/d\omega$ as

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

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

$$
\left(\frac{dD}{d\omega}\right)_{dy=db=0} = \frac{\left(\Theta - h\right)\left(\gamma f_1'\omega - v''h\right) p\left(q_c\right)\gamma f_{12}''}{q_c\Delta} > 0. \tag{A27}
$$

Finally, assume $dw = d\omega = 0$. From (A18)-(A19) 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)

$$
\begin{aligned}
\left(\frac{dh}{d\gamma}\right)_{dy=db=0} &= \frac{\omega f_1'\left[\left(1 - \beta\right) p''\left(q_c\right) - \left(\Theta - h\right)\gamma f_{22}''\right] + f_2'\gamma\left(\Theta - h\right)\omega f_{12}'' - q_c f_2'\left(1 - \beta\right) p''\left(q_c\right)}{\Delta} \\
&\quad - \left[\frac{\left(\omega f_1' - q_c f_2'\right)}{\Delta} p'\left(q_c\right) + \frac{f_2'}{\Delta} p\left(q_c\right)\right]\left(1 - \beta\right)^2\left(\Theta - h\right) u''\left(c\right) p'\left(q_c\right), \tag{A28}
\end{aligned}
$$

$$
\begin{aligned}
\left(\frac{dq_c}{d\gamma}\right)_{dy=db=0} &= \frac{\left(\omega f_{12}'' - q_c f_{22}''\right)\omega f_1' + \left[\left(f_{12}'' q_c - \omega f_{11}''\right)\omega - \frac{v''}{\gamma}\right] f_2'}{\Delta}\gamma \\
&\quad - \left[\frac{f_2'}{\Delta} p\left(q_c\right) + \frac{\left(\omega f_1' - q_c f_2'\right)}{\Delta} p'\left(q_c\right)\right]\left(1 - \beta\right)^2 u''\left(c\right) p\left(q_c\right). \tag{A29}
\end{aligned}
$$

From (A28)-(A29) we can calculate $dD/d\gamma$ as

$$
\begin{aligned}
\left(\frac{dD}{d\gamma}\right)_{dy=db=0} &= -p\left(q_c\right)\frac{dh}{d\gamma} + \left(\Theta - h\right) p'\left(q_c\right)\frac{dq_c}{d\gamma} \\
&= \frac{q_c f_2' - \omega f_1'}{\Delta} p\left(q_c\right)\left(1 - \beta\right) p''\left(q_c\right) u'\left(c\right) \\
&\quad + \frac{\left(\Theta - h\right) p\left(q_c\right)}{\Delta}\left(f_1' f_{22}'' - f_2' f_{12}''\right)\left(1 - \epsilon_{p,q_c}\right)\omega\gamma \\
&\quad + \frac{\left(\Theta - h\right) p\left(q_c\right)}{\Delta}\left[\left(\omega\right)^2\left(f_1' f_{12}'' - f_2' f_{11}''\right)\gamma - v'' f_2'\right]\frac{\epsilon_{p,q_c}}{q_c}. \tag{A30}
\end{aligned}
$$

9

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 (A24), (A27) and (A30) 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 (A15) 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 (A23) 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 (A15), child care expenditures should optimally be subsidized ($\beta > 0$). $\square$

Finally, Corollary 3 provides an example of an optimum where child care expenditures should be taxed and all agents face a positive marginal income tax rate.

Corollary 2. *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 Assume that $u'' = 0$, $p'' = 0$, $f_{12}'' > 0$, $\gamma^1 = \gamma^2$ and $\omega^1 = \omega^2$. From (A24) we already know that $dD/dw > 0$, which implies $\beta < 0$, i.e. a proportional tax on child care expenditures.

From (A9), and taking into account that $u'' = 0 \implies dD/db = 0$, we have

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

Noticing that

$$\frac{dD}{dy} = -\left(\frac{dD}{dw}\right)_{dy=0} \frac{w}{y}, \tag{A31}$$

from (A24) we have that $p'' = 0$ implies

$$\frac{dD}{dy} = \frac{(\Theta - h) \, p' \, (q_c) \, \gamma \omega f_{12}''}{\Delta} \frac{1}{w} v'' < 0. \tag{A32}$$

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

From (A10), and taking into account that $u'' = 0$ implies $dD/db = 0$ and also $\widehat{V}_b = V_b^1$, we have

$$T' \left(y^1\right) = \frac{\lambda}{\mu \pi} \left(\widehat{V}_y - V_y^1\right) + \frac{dD^1}{dy^1} \beta. \tag{A33}$$

From (A32) 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

$$\Theta - \frac{y^1}{w^2} - \widehat{h} > \Theta - \frac{y^1}{w^1} - h^1. \tag{A34}$$

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 (A21) 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 (A20), 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:

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

Concavity of the $f(\cdot,\cdot)$-function ($f_{11}'' f_{22}'' - (f_{12}'')^2 > 0$) implies $1 - \frac{(\Theta - h)\gamma f_{22}'' v''}{\Delta} > 0$, 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 (A34) is satis-

fied. Based on this, we can conclude that the first term appearing on the right hand side of (A33) is positive too, and therefore $T'(y^1) > 0$. $\square$

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 (A1) with $\beta^2 = 0$) that high-skilled face a zero marginal income tax rate. The following Corollary summarizes the results for this case.

Corollary 3. *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 (A2), 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 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

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

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

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

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 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:

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

$$(\delta + \lambda) V_b^2 = \mu (1 - \pi) \left\{ 1 + \beta^2 \left[\left(\Theta - h^2 \right) p' \left(q_c^2 \right) \frac{\partial q_c^2}{\partial b^2} - p \left(q_c^2 \right) \frac{\partial h^2}{\partial b^2} \right] \right\}, \quad \text{(A35)}$$

$$(\delta + \lambda) V_{\beta^2}^2 - \mu (1 - \pi) \left[D^2 + \beta^2 \frac{\partial D^2}{\partial \beta^2} \right] = 0, \quad \text{(A36)}$$

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

$$V_b^1 = \lambda \widehat{V}_b^2 + \mu \pi \left\{ 1 + \beta^1 \left[\left(\Theta - h^1 \right) p' \left(q_c^1 \right) \frac{\partial q_c^1}{\partial b^1} - p \left(q_c^1 \right) \frac{\partial h^1}{\partial b^1} \right] \right\}, \quad \text{(A37)}$$

$$V_{\beta^1}^1 - \lambda \widehat{V}_{\beta^1} - \mu \pi \left[D^1 + \beta^1 \frac{\partial D^1}{\partial \beta} \right] = 0. \quad \text{(A38)}$$

Applying Roy's identity to (A36) 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 (A35) 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:

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

Applying Roy's identity to (A38) 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 (A37) 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:

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

□

B Proofs and derivations

B.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:

$$(\delta + \lambda) \frac{v'\left(\Theta - \frac{y^2}{w^2} - h^2\right)}{w^2} = \mu(1 - \pi) \tag{B1}$$

$$(\delta + \lambda) u'\left(c^2\right) = \mu(1 - \pi) \tag{B2}$$

$$(\delta + \lambda)\left\{\gamma^2\left[\omega^2 f_1'\left(\omega^2 h^2, \left(\Theta - h^2\right) q_c^2\right) - q_c^2 f_2'\left(\omega^2 h^2, \left(\Theta - h^2\right) q_c^2\right)\right] - v'\left(\Theta - \frac{y^2}{w^2} - h^2\right)\right\}$$
$$= -\mu(1 - \pi) p\left(q_c^2\right) \tag{B3}$$

$$(\delta + \lambda)\gamma^2 f_2'\left(\omega^2 h^2, \left(\Theta - h^2\right) q_c^2\right) = \mu(1 - \pi) p'\left(q_c^2\right) \tag{B4}$$

Using (B2) and taking into account that $\ell^2 = \Theta - \frac{y^2}{w^2} - h^2$, one can rewrite conditions (B1), (B3) and (B4) as respectively:

$$1 - \frac{v'\left(\ell^2\right)}{w^2 u'\left(c^2\right)} = 0, \tag{B5}$$

$$p\left(q_c^2\right) + \frac{\gamma^2\left[\omega^2 f_1'\left(\omega^2 h^2, \left(\Theta - h^2\right) q_c^2\right) - q_c^2 f_2'\left(\omega^2 h^2, \left(\Theta - h^2\right) q_c^2\right)\right]}{u'\left(c^2\right)} - \frac{v'\left(\ell^2\right)}{u'\left(c^2\right)} = 0, \tag{B6}$$

$$\frac{\gamma^2 f_2'\left(\omega^2 h^2, \left(\Theta - h^2\right) q_c^2\right)}{u'\left(c^2\right)} - p'\left(q_c^2\right) = 0, \tag{B7}$$

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:

$$\frac{v'\left(\Theta - \frac{y^1}{w^1} - h^1\right)}{w^1} = \lambda \frac{v'\left(\Theta - \frac{y^1}{w^2} - h^1\right)}{w^2} + \mu\pi \tag{B8}$$

$$u'\left(c^1\right) = \lambda u'\left(c^1\right) + \mu\pi \tag{B9}$$

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

$$= \lambda \left\{\gamma^2 \left[\omega^2 f_1'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right) - q_c^1 f_2'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right)\right] - v'\left(\Theta - \frac{y^1}{w^2} - h^1\right)\right\}$$

$$- \mu\pi p\left(q_c^1\right) \tag{B10}$$

$$\gamma^1 f_2'\left(\omega^1 h^1, \left(\Theta - h^1\right) q_c^1\right) = \lambda \gamma^2 f_2'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right) + \mu\pi p'\left(q_c^1\right) \tag{B11}$$

Combining (B8) and (B9) 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

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

Combining (B10) and (B9) one gets:

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

$$= \frac{\lambda}{\mu\pi}\left\{\gamma^2 \left[\omega^2 f_1'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right) - q_c^1 f_2'\left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right)\right] - v'\left(\Theta - \frac{y^1}{w^2} - h^1\right)\right\}$$

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

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

$$\gamma^2 \left[\omega^2 f_1' \left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right) - q_c^1 f_2' \left(\omega^2 h^1, \left(\Theta - h^1\right) q_c^1\right)\right] - v' \left(\Theta - \frac{y^1}{w^2} - h^1\right)$$

$$>$$

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

$$(B12)$$

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 (B12) 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).[3]

Finally, combining (B11) and (B9) 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.

---

[3]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:

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

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

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

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

B.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:

$$(\delta + \lambda) \frac{v'\left(\frac{D^2}{p(q_c^2)} - \frac{y^2}{w^2}\right)}{w^2} = \mu(1 - \pi), \tag{B13}$$

$$(\delta + \lambda) u'\left(c^2\right) = \mu(1 - \pi), \tag{B14}$$

$$\frac{\delta + \lambda}{p(q_c^2)} \left[-\omega^2 f_1'\left(\left(\Theta - \frac{D^2}{p(q_c^2)}\right)\omega^2, \frac{D^2 q_c^2}{p(q_c^2)}\right) + q_c^2 f_2'\left(\left(\Theta - \frac{D^2}{p(q_c^2)}\right)\omega^2, \frac{D q_c^2}{p(q_c^2)}\right) \right] \gamma^2$$
$$+ \frac{\delta + \lambda}{p(q_c^2)} v'\left(\frac{D^2}{p(q_c^2)} - \frac{y^2}{w^2}\right)$$
$$= \mu(1 - \pi). \tag{B15}$$

Using (B14) and taking into account that $\ell^2 = \frac{D^2}{p(q_c^2)} - \frac{y^2}{w^2}$, one can rewrite conditions (B13) and (B15) as, respectively:

$$1 - \frac{v'(\ell^2)}{w^2 u'(c^2)} = 0, \tag{B16}$$

$$p\left(q_c^2\right) + \frac{\left[\omega^2 f_1'\left(\left(\Theta - \frac{D^2}{p(q_c^2)}\right)\omega^2, \frac{D^2 q_c^2}{p(q_c^2)}\right) - q_c^2 f_2'\left(\left(\Theta - \frac{D^2}{p(q_c^2)}\right)\omega^2, \frac{D q_c^2}{p(q_c^2)}\right)\right] \gamma^2}{u'(c^2)} - \frac{v'(\ell^2)}{u'(c^2)} = 0. \tag{B17}$$

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[4]

$$\left[\omega^2 f_1'\left(\left(\Theta - \frac{D^2}{p(q_c^2)}\right)\omega^2, \frac{D^2 q_c^2}{p(q_c^2)}\right) + \left(\frac{p(q_c^2)}{p'(q_c^2)} - q_c^2\right) f_2'\left(\left(\Theta - \frac{D^2}{p(q_c^2)}\right)\omega^2, \frac{D^2 q_c^2}{p(q_c^2)}\right)\right] \gamma^2 = v'(\ell^2), \tag{B18}$$

---

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

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

The associated first order condition is

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

17

by combining (B17) and (B18) one gets

$$\gamma^2 q_c^2 f_2' \left(\left(\Theta - \frac{D^2}{p\left(q_c^2\right)} \right) \omega^2, \frac{D^2 q_c^2}{p\left(q_c^2\right)} \right) = p'\left(q_c^2\right) u'\left(c^2\right). \tag{B19}$$

Comparing (B16), (B17) and (B19) 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:

$$\frac{v'\left(\frac{D^1}{p(q_c^1)} - \frac{y^1}{w^1}\right)}{w^1} = \lambda \frac{v'\left(\frac{D^1}{p(\widehat{q}_c)} - \frac{y^1}{w^2}\right)}{w^2} + \mu\pi, \tag{B20}$$

$$u'\left(c^1\right) = \lambda u'\left(c^1\right) + \mu\pi, \tag{B21}$$

$$\begin{aligned}
& \left[-\omega^1 f_1' \left(\left(\Theta - \frac{D^1}{p\left(q_c^1\right)} \right) \omega^1, \frac{D^1 q_c^1}{p\left(q_c^1\right)} \right) + q_c^1 f_2' \left(\left(\Theta - \frac{D^1}{p\left(q_c^1\right)} \right) \omega^1, \frac{D^1 q_c^1}{p\left(q_c^1\right)} \right) \right] \frac{\gamma^1}{p\left(q_c^1\right)} \\
& + \frac{v'\left(\frac{D^1}{p(q_c^1)} - \frac{y^1}{w^1}\right)}{p\left(q_c^1\right)} \\
= & \frac{\lambda \gamma^2}{p\left(\widehat{q}_c\right)} \left[-\omega^2 f_1' \left(\left(\Theta - \frac{D^1}{p\left(\widehat{q}_c\right)} \right) \omega^2, \frac{D^1 \widehat{q}_c}{p\left(\widehat{q}_c\right)} \right) + \widehat{q}_c f_2' \left(\left(\Theta - \frac{D^1}{p\left(\widehat{q}_c\right)} \right) \omega^2, \frac{D^1 \widehat{q}_c}{p\left(\widehat{q}_c\right)} \right) \right] \\
& + \frac{\lambda}{p\left(\widehat{q}_c\right)} v'\left(\frac{D^1}{p\left(\widehat{q}_c^2\right)} - \frac{y^1}{w^2}\right) + \mu\pi. \tag{B22}
\end{aligned}$$

Combining (B20) and (B21) gives:

$$1 - \frac{v'\left(\frac{D^1}{p(q_c^1)} - \frac{y^1}{w^1}\right)}{w^1 u'\left(c^1\right)} = \frac{\lambda}{\mu\pi} \left[\frac{v'\left(\frac{D^1}{p(q_c^1)} - \frac{y^1}{w^1}\right)}{w^1} - \frac{v'\left(\frac{D^1}{p(\widehat{q}_c)} - \frac{y^1}{w^2}\right)}{w^2} \right]. \tag{B23}$$

Combining (B22) and (B21) gives

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

18

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

$$\left[\left(-\omega^2 f_1'\left(\left(\Theta - \frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2, \frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right) + \widehat{q}_c f_2'\left(\left(\Theta - \frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2, \frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)\right)\gamma^2\right]\frac{1}{p\left(\widehat{q}_c\right)}$$

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

$$>$$

$$\left[\left(-\omega^1 f_1'\left(\left(\Theta - \frac{D^1}{p\left(q_c^1\right)}\right)\omega^1, \frac{D^1 q_c^1}{p\left(q_c^1\right)}\right) + q_c^1 f_2'\left(\left(\Theta - \frac{D^1}{p\left(q_c^1\right)}\right)\omega^1, \frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)\right)\gamma^1\right]\frac{1}{p\left(q_c^1\right)}$$

$$+\frac{v'\left(\frac{D^1}{p\left(q_c^1\right)} - \frac{y^1}{w^1}\right)}{p\left(q_c^1\right)}. \tag{B24}$$

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:

$$\frac{p\left(q_c^1\right)}{p'\left(q_c^1\right)}\gamma^1 f_2'\left(\left(\Theta - \frac{D^1}{p\left(q_c^1\right)}\right)\omega^1, \frac{D^1 q_c^1}{p\left(q_c^1\right)}\right) = v'\left(\frac{D^1}{p\left(q_c^1\right)} - \frac{y^1}{w^1}\right)$$

$$+q_c^1 f_2'\left(\left(\Theta - \frac{D^1}{p\left(q_c^1\right)}\right)\omega^1, \frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)\gamma^1$$

$$-\omega^1 f_1'\left(\left(\Theta - \frac{D^1}{p\left(q_c^1\right)}\right)\omega^1, \frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)\gamma^1; \tag{B25}$$

$$\frac{p\left(\widehat{q}_c\right)}{p'\left(\widehat{q}_c\right)}\gamma^2 f_2'\left(\left(\Theta - \frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2, \frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right) = v'\left(\frac{D^1}{p\left(\widehat{q}_c\right)} - \frac{y^1}{w^2}\right)$$

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

$$-\left(\omega^2 f_1'\left(\left(\Theta - \frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2, \frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)\right)\gamma^2. \tag{B26}$$

Substituting (B25)-(B26) into (B24) and simplifying terms, one obtains that there should be a downward distortion on D^1 iff

$$\frac{\gamma^2 f_2'\left(\left(\Theta - \frac{D^1}{p\left(\widehat{q}_c\right)}\right)\omega^2, \frac{D^1\widehat{q}_c}{p\left(\widehat{q}_c\right)}\right)}{p'\left(\widehat{q}_c\right)} > \frac{\gamma^1 f_2'\left(\left(\Theta - \frac{D^1}{p\left(q_c^1\right)}\right)\omega^1, \frac{D^1 q_c^1}{p\left(q_c^1\right)}\right)}{p'\left(q_c^1\right)}. \tag{B27}$$

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:

$$\frac{\left(p'\left(q_c\right)\right)^2 - p\left(q_c\right)p''\left(q_c\right)}{\left(p'\left(q_c\right)\right)^2}\gamma f_2' dq_c + \frac{p\left(q_c\right)}{p'\left(q_c\right)}\gamma\left\{\omega\frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}f_{12}'' + \left(\frac{D}{p\left(q_c\right)} - \frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}q_c\right)f_{22}''\right\}dq_c$$

$$+\gamma\left\{\omega\frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}f_{11}'' + \left(\frac{D}{p\left(q_c\right)} - \frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}q_c\right)f_{12}''\right\}\omega dq_c$$

$$-\gamma\left\{\omega\frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}f_{12}'' + \left(\frac{D}{p\left(q_c\right)} - \frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}q_c\right)f_{22}''\right\}q_c dq_c$$

$$+\frac{Dp'\left(q_c\right)}{\left(p\left(q_c\right)\right)^2}v''dq_c - \gamma f_2' dq_c - \frac{y}{\left(w\right)^2}v''dw$$

$$+\gamma f_1' d\omega + \left(\Theta - \frac{D}{p\left(q_c\right)}\right)\gamma\omega f_{11}'' d\omega + \frac{p\left(q_c\right)}{p'\left(q_c\right)}\left(\Theta - \frac{D}{p\left(q_c\right)}\right)\gamma f_{12}'' d\omega - \left(\Theta - \frac{D}{p\left(q_c\right)}\right)\gamma q_c f_{12}'' d\omega$$

$$+\omega f_1' d\gamma + \left(\frac{p\left(q_c\right)}{p'\left(q_c\right)} - q_c\right)f_2' d\gamma$$

$$= 0.$$

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

$$\Lambda \equiv \left\{\left[\gamma\left(\omega f_{11}'' - q_c f_{12}''\right)\omega + v''\right]\frac{p'\left(q_c\right)}{p\left(q_c\right)} + \left[\left(\frac{1}{\epsilon_{p,q_c}} - 1\right)q_c f_{22}'' + \omega f_{12}''\right]\left(1 - \epsilon_{p,q_c}\right)\gamma\right\}\frac{D}{p\left(q_c\right)}$$

$$+\left[\omega\frac{D}{p\left(q_c\right)}f_{12}'' - \frac{p\left(q_c\right)p''\left(q_c\right)}{\left(p'\left(q_c\right)\right)^2}f_2'\right]\gamma, \tag{B28}$$

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

$$\left(\frac{dq_c}{dw}\right)_{dD=0,dy=0} = \frac{yv''}{\left(w\right)^2 \Lambda} > 0, \tag{B29}$$

$$\left(\frac{dq_c}{d\omega}\right)_{dD=0,dy=0} = -\frac{f_1' + \left(\Theta - \frac{D}{p(q_c)}\right)\left[\omega f_{11}'' + \left(\frac{1}{\epsilon_{p,q_c}} - 1\right)q_c f_{12}''\right]}{\Lambda}\gamma, \tag{B30}$$

$$\left(\frac{dq_c}{d\gamma}\right)_{dD=0,dy=0} = -\frac{\omega f_1' + \left(\frac{1}{\epsilon_{p,q_c}} - 1\right)q_c f_2'}{\Lambda}. \tag{B31}$$

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 $(dq_c/d\omega)_{dD=0,dY=0}$ as

$$\left(\frac{dq_c}{d\omega}\right)_{dD=0,dy=0} = -\frac{\left(1 + \epsilon_{f'_{1,\omega}}\right) f'_1 + \left(\Theta - \frac{D}{p(q_c)}\right)\left(\frac{1}{\epsilon_{p,q_c}} - 1\right) q_c f''_{12}}{\Lambda}\gamma. \qquad (B32)$$

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 (B29) 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 (B28), we have:

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

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 $p(\cdot)$,[5] 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 (B23) is strictly positive, which in turn implies that y^1 is downward distorted $\left(1 - v'\left(\frac{D^1}{p(q_c^1)} - \frac{y^1}{w^1}\right)/w^1 u'(c^1) > 0\right)$. 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 (B23) is negative, implying that y^1 ought to be distorted upwards.

Consider now the condition determining whether a downward distortion on D^1 is

---

[5]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 $0 < 1 - \frac{\epsilon_{p',q_c}}{\epsilon_{p,q_c}} \leq 1$.

optimal, i.e. condition (B27). We have that, keeping fixed D,

$$\left(\frac{d\left(\frac{\gamma f_2'}{p'(q_c)}\right)}{dw}\right)_{dD=0} = \left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0} \left(\frac{dq_c}{dw}\right)_{dD=dy=0}, \tag{B33}$$

$$\left(\frac{d\left(\frac{\gamma f_2'}{p'(q_c)}\right)}{d\omega}\right)_{dD=0} = \left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0} \left(\frac{dq_c}{d\omega}\right)_{dD=dy=0} + \left(\Theta - \frac{D}{p(q_c)}\right)\frac{\gamma f_{12}''}{p'(q_c)}, \tag{B34}$$

$$\left(\frac{d\left(\frac{\gamma f_2'}{p'(q_c)}\right)}{d\gamma}\right)_{dD=0} = \left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0} \left(\frac{dq_c}{d\gamma}\right)_{dD=dy=0} + \frac{f_2'}{p'(q_c)}. \tag{B35}$$

Moreover,

$$\left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0} = \frac{\left[\frac{p'(q_c)}{p(q_c)}\omega f_{12}'' + \left(1 - \frac{p'(q_c)}{p(q_c)}q_c\right)f_{22}''\right]\frac{Dp'(q_c)}{p(q_c)} - f_2'p''(q_c)}{[p'(q_c)]^2}\gamma, \tag{B36}$$

and therefore:

$$sign\left\{\left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0}\right\} = sign\left\{\frac{p'(q_c)}{p(q_c)}\omega f_{12}'' + \left(1 - \frac{p'(q_c)}{p(q_c)}q_c\right)f_{22}'' - f_2'p''(q_c)\frac{p(q_c)}{Dp'(q_c)}\right\}$$

$$= sign\left\{\frac{p'(q_c)}{p(q_c)}\omega f_{12}'' + \left(1 - \frac{p'(q_c)}{p(q_c)}q_c\right)f_{22}'' - f_2'\frac{p''(q_c)}{p'(q_c)}\frac{1}{h_c}\right\}$$

$$= sign\left\{\left(\epsilon_{p,q_c}\omega f_{12}'' - \epsilon_{p',q_c}\frac{f_2'}{h_c}\right)\frac{1}{q_c} + (1 - \epsilon_{p,q_c})f_{22}''\right\}. \tag{B37}$$

B.3   Proof of Corollary 1

Part i). Assume, as in Corollary 1, that $f_{12}'' > 0$, $p'' = 0$ (so that $p(q_c) = kq_c$, $\epsilon_{p,q_c} = 1$ and $\epsilon_{p',q_c} = 0$), $\gamma^1 \leq \gamma^2$ and $\omega^1 \leq \omega^2$. It follows from (B37) 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 (B29) we have that

$\left(\frac{dq_c}{dw}\right)_{dD=dy=0} > 0$ and from (B31) we have that $\left(\frac{dq_c}{d\gamma}\right)_{dD=dy=0} > 0$. Thus, from (B33) and

(B35) 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

(B27) 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:

$$\left(\frac{dq_c}{d\omega}\right)_{dD=0,dy=0} = -\frac{\left(1 + \epsilon_{f_1',\omega}\right)f_1'}{\Lambda}\gamma, \tag{B38}$$

where Λ, defined in (B28), simplifies to:

$$\Lambda = \left[(\omega)^2 \gamma f_{11}'' + v''\right] \frac{p'(q_c) D}{(p(q_c))^2}. \tag{B39}$$

Moreover, with $\epsilon_{p,q_c} = 1$, (B36) simplifies to:

$$\left(\frac{\partial \frac{\gamma f_2'}{p'(q_c)}}{\partial q_c}\right)_{dD=0} = \frac{\omega f_{12}'' D}{[p(q_c)]^2} \gamma. \tag{B40}$$

Therefore, substituting (B38)-(B40) into (B34) gives:

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

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 (B37) 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 (B29) we have that $\left(\frac{dq_c}{dw}\right)_{dD=dY=0} > 0$, we can conclude from (B33) that condition (B27) does not hold, implying that D^1 should optimally be upward distorted.

B.4   Derivation of (11) and (12)

From condition (9) one obtains:

$$db^1 = -\left(\Theta - h^{1,in}\right) p(\bar{q}_c) d\beta - \frac{[\gamma f_2' - (1 - \beta) p' u'] (\Theta - h^{1,in})}{u'} d\bar{q}_c. \tag{B41}$$

If we now substitute (B41) in (10), this gives:

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

23

or, rearranging terms:

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

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

$$+ \gamma f_{22} \left(\Theta - h^{1,in} \right) d\overline{q}_c$$

$$= \quad 0. \tag{B42}$$

Totally differentiating the first order condition (8), that applies to low-skilled households who opt-in, gives:

$$\frac{dh^{1,in}}{d\beta} = \frac{u'}{v'' + \gamma \left[(\omega)^2 f_{11}'' - 2\omega \overline{q}_c f_{12}'' + (\overline{q}_c)^2 f_{22}'' \right]} p \left(\overline{q}_c \right), \tag{B43}$$

$$\frac{dh^{1,in}}{d\overline{q}_c} = \frac{\left[f_2' - \left(\Theta - h^{1,in} \right) \left(\omega f_{12}'' - \overline{q}_c f_{22}'' \right) \right] \gamma - (1-\beta) p'u'}{v'' + \gamma \left[(\omega)^2 f_{11}'' - 2\omega \overline{q}_c f_{12}'' + (\overline{q}_c)^2 f_{22}'' \right]}, \tag{B44}$$

$$\frac{dh^{1,in}}{db^1} = \quad 0. \tag{B45}$$

Substituting (B43)-(B45) in (B42) gives:

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

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

$$+ \gamma f_{22} \left(\Theta - h^{1,in} \right) d\overline{q}_c = 0,$$

from which one obtains

$$d\overline{q}_c = \frac{\left(\omega f_{12}'' - \overline{q}_c f_{22}'' \right) p \left(\overline{q}_c \right) u'}{\left\{ \gamma (\omega)^2 \left[(f_{12}'')^2 - f_{11}'' f_{22}'' \right] - v'' f_{22} \right\} \left(\Theta - h^{1,in} \right) - \left(\omega f_{12}'' - \overline{q}_c f_{22}'' \right) \left[\gamma f_2' - (1-\beta) p'u' \right]} d\beta. \tag{B46}$$

Substituting the value found above for $d\overline{q}_c$ into (B41) gives:

$$db^1 = - \left\{ \frac{\gamma (\omega)^2 \left[(f_{12}'')^2 - f_{11}'' f_{22}'' \right] - v'' f_{22}''}{\gamma (\omega)^2 \left[(f_{12}'')^2 - f_{11}'' f_{22}'' \right] - v'' f_{22}'' - \frac{\left(\omega f_{12}'' - \overline{q}_c f_{22}'' \right) \left[\gamma f_2' - (1-\beta) p'(\overline{q}_c)u' \right]}{\Theta - h^{1,in}}} \right\} \left(\Theta - h^{1,in} \right) p \left(\overline{q}_c \right) d\beta. \tag{B47}$$

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. (B46)-(B47) can be simplified to obtain (11)-(12).

B.5 Derivation of the negative welfare effect of an opting-out public provision scheme on the utility of mimicking households

With respect to the impact on the first self-selection constraint ($V^2\left(b^2, y^2\right) \geq V^2\left(b^1, y^1\right)$) we have a positive mimicking-deterring effect since mimickers' utility change by

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

With respect to the impact on the second self-selection constraint ($V^2\left(b^2, y^2\right) \geq V^2\left(\bar{q}_c, \beta, b^1, y^1\right)$) we have that mimickers' utility change by

$$
\begin{aligned}
&dV^2\left(\bar{q}_c, \beta, b^1, y^1\right)\\
=\;& \frac{\partial V^2\left(\bar{q}_c, \beta, b^1, y^1\right)}{\partial \bar{q}_c}d\bar{q}_c + \frac{\partial V^2\left(\bar{q}_c, \beta, b^1, y^1\right)}{\partial \beta}d\beta + \frac{\partial V^2\left(\bar{q}_c, \beta, b^1, y^1\right)}{\partial b^1}db^1\\
=\;& \left(\Theta - \widehat{h}^{in}\right)\gamma\widehat{f}_2^{\prime in}\left(\omega\widehat{h}^{in}, \left(\Theta - \widehat{h}^{in}\right)\bar{q}_c\right)\frac{\left(\omega f_{12}'' - \bar{q}_c f_{22}''\right)p\left(\bar{q}_c\right)u'}{\gamma\left(\Theta - h^{1,in}\right)\left(\omega\right)^2\left[\left(f_{12}''\right)^2 - f_{11}''f_{22}''\right] - \left(\Theta - h^{1,in}\right)v''f_{22}''}d\beta\\
&- \left(\Theta - \widehat{h}^{in}\right)\left(1 - \beta\right)p'\left(\bar{q}_c\right)u'\frac{\left(\omega f_{12}'' - \bar{q}_c f_{22}''\right)p\left(\bar{q}_c\right)u'}{\gamma\left(\Theta - h^{1,in}\right)\left(\omega\right)^2\left[\left(f_{12}''\right)^2 - f_{11}''f_{22}''\right] - \left(\Theta - h^{1,in}\right)v''f_{22}''}d\beta\\
&+ p\left(\bar{q}_c\right)\left(\Theta - \widehat{h}^{in}\right)u'd\beta - \left(\Theta - h^{1,in}\right)p\left(\bar{q}_c\right)u'd\beta\\
=\;& \left[\gamma\widehat{f}_2^{\prime in}\left(\omega\widehat{h}^{in}, \left(\Theta - \widehat{h}^{in}\right)\bar{q}_c\right) - \left(1 - \beta\right)p'\left(\bar{q}_c\right)u'\right]\frac{\Theta - \widehat{h}^{in}}{\Theta - h^{1,in}}\frac{\left(\omega f_{12}'' - \bar{q}_c f_{22}''\right)p\left(\bar{q}_c\right)u'}{\gamma\left(\omega\right)^2\left[\left(f_{12}''\right)^2 - f_{11}''f_{22}''\right] - v''f_{22}''}d\beta\\
&+ p\left(\bar{q}_c\right)\left(h^{1,in} - \widehat{h}^{in}\right)u'd\beta. \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (B48)
\end{aligned}
$$

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:

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

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

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

Thus, defining Υ as

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

we have:

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

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

Thus, the last term on the right hand side of (B48) 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 $\bar{q}_c$ satisfied $p'(\bar{q}_c) = \frac{\gamma f_2'(\omega h^{1,in}, (\Theta - h^{1,in})\bar{q}_c)}{u'}$, having established that $h^{1,in} - \hat{h}^{in} < 0$ allows concluding that $\gamma \hat{f}_2'^{in} \left(\omega\hat{h}^{in}, \left(\Theta - \hat{h}^{in} \right) \bar{q}_c \right) - (1 - \beta) p'(\bar{q}_c) u' > 0$. We can then conclude that the proposed reform also has a detrimental effect on a high-skilled households who mimic and opt in.

C Child care subsidies in the United States

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 1.

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

Table 1: 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%		

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 1.

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).

D 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.[6] Given the large number of private decision variables, we did not find a procedure that incorporates the first-order conditions as constraints to be very robust. Instead, we compute the solutions to the individual decision problems numerically using a nested optimization procedure. In contrast to the first-order approach, this procedure allows us to take into account both first and second order conditions in the individual optimization problem. The drawback is that we have to rely on numerical approximations of derivatives in the upper level which significantly increases the time it takes to find an optimal solution. In addition there is a computational overhead associated with the nested optimization layer. To increase performance, exact first and second order derivatives to the lower level optimization problem were provided to the numerical optimization algorithm and we relied on a fast 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.[7]

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

---

[6]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.

[7]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.

agent we introduce in the economy, we need to expand the set of pre-tax/post-tax income points offered by the government, which increases the number of control variables that need to be optimized in the "main" government problem. These additional income points also generate additional self-selection constraints, making it more difficult to achieve convergence in the main problem. Finally, and perhaps most critically, as explained below, adding types increases the number of marginal workers that need to be identified in order to determine the number of mothers who find it optimal to work.

There are two approaches to modeling the extensive margin. One approach is to let agents optimally choose their labor force participation status in the lower level optimization problem. This implies that the fraction of workers at each skill level is endogenous to the tax system. While this does not introduce any non-smoothness in the government's social welfare function or tax revenue function (provided the number of cost types is sufficiently large), it does imply that individuals might switch discretely from working to not working, or vice versa, in response to a small change in the income tax. This causes an undesirable reshaping of the set of incentive constraints, which makes it difficult to find solutions to the government's problem using gradient-based optimization algorithms. We have therefore refrained from this approach. Instead, we add the binary variables associated with mothers' labor force participation decision as artificial control variables of the government, while adding a set of constraints ensuring that the labor force participation decisions assigned to agents are incentive-compatible. The benefit of this approach is that the marginal control variables can be treated as exogenous and optimized in a separate optimization layer. This means that our optimization problem has three layers. An outer layer where we choose the labor force participation levels at each skill level (equivalent to identifying the marginal worker), a middle layer where we choose the pre-tax/post-tax income 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.[8]

E   Robustness with respect to specification of innate ability

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

---

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

Table 2: 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).

References

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

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