DISCUSSION PAPER SERIES

IZA DP No. 12718

Having It All, for All: Child-Care Subsidies and Income Distribution Reconciled

Francesca Barigozzi
Helmuth Cremer
Kerstin Roeder

OCTOBER 2019

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12718

Having It All, for All: Child-Care Subsidies and Income Distribution Reconciled

Francesca Barigozzi
University of Bologna

Helmuth Cremer
Toulouse School of Economics and IZA

Kerstin Roeder
University of Augsburg

OCTOBER 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Having It All, for All: Child-Care Subsidies and Income Distribution Reconciled*

We study the design of child-care policies when redistribution matters. Traditional mothers provide some informal child care, whereas career mothers purchase full time formal care. The sorting of women across career paths is endogenous and shaped by a social norm about gender roles in the family. Via this social norm traditional mothers' informal child care imposes an externality on career mothers, so that the market outcome is inefficient. Informal care is too large and the group of career mothers is too small so that inefficiency and gender inequality go hand in hand. In a first-best world redistribution across couples and efficiency are separable. Redistribution is performed via lump-sum transfers and taxes which are designed to equalize utilities across all couples. The efficient allocation of child care is obtained by subsidizing formal care at a Pigouvian rate. However, in a second-best setting, a trade-off between efficiency and redistribution emerges. The optimal uniform subsidy is lower than the "Pigouvian" level. Conversely, under a non-linear policy the first-best "Pigouvian" rule for the (marginal) subsidy on informal care is reestablished. While the share of high career mothers continues to be distorted downward for incentive reasons, this policy is effective in reconciling the objectives of reducing the child care related inefficiency and achieving a more equal income distribution across couples. Our results continue to hold when the social norm is defined within the mothers' social group, rather than being based on the entire population.

JEL Classification: D13, H23, J16, J22

Keywords: social norms, child-care, women's career choices, child care subsidies, redistribution of income

Corresponding author:
Helmuth Cremer
Toulouse School of Economics
21 Allée de Brienne
31015 Toulouse
France
E-mail: helmuth.cremer@tse-fr.eu

* Financial support from the Chaire "Marché des risques et creation de valeur" of the FdR/SCOR is gratefully acknowledged. Helmuth Cremer acknowledges funding received by TSE from ANR under grant ANR-17-EURE- 0010 (Investissements d.Avenir program).

1 Introduction

While female labor force participation has been increasing steadily over the last decades (Goldin, 2006 and 2014b, Kleven and Landais, 2017) mothers continue to be the main providers of child care within the family (e.g., Paull, 2008; Ciccia and Verloo, 2012). Maternity leave and other child related career breaks or part-time work contribute in a significant way to the persistence of gender inequalities in the labor market. The so called "child penalty" appears to explain up to about 80% of the gender wage gap; see Kleven *et al.* (2018).

As a possible reason for the persistence of child-care compatible (part-time) work and child penalties, many studies point to social norms shaping women's preferences over family and career (see Fortin 2015, Farré and Vella 2013, Bertrand *et al.* 2015, Bursztyn *et al.* 2017 and Kleven *et al.* 2018, among others). Social norms may contribute to the differential sorting of men and women across occupations with women entering low pay occupations that allow for shorter working days or more flexible working hours (see Goldin, 2014 and Card *et al.* 2016).

During the last five decades, most developed countries have put into practice multiple child policies with various declared goals, including gender equity, higher fertility, and child development. The policies who seem to have been the most effective in reducing gender disparities are child care provision and subsidization. Evidence indicates that early childhood spending contributed substantially to enabling women to combine working life and motherhood, and to altering social norms regarding gender roles; see Olivetti and Petrongolo (2017).

Reducing gender disparities in the labor market is not the unique concern which is relevant for child care policies. Redistribution across income levels has been *the* major issue for the design of tax and expenditures policies, it has lead to the emergence of the concept of "welfare state" which applies to all developed countries albeit to a different degree; see Boadway and Keen (1993). Unfortunately, the objective of reducing gender disparities in the labor market and redistributive concerns may be conflicting goals. Specifically, child care provision and subsidization may be regressive if the parents who benefit more from the policy are the ones with relatively higher income. This seems to be the case in most OECD countries, where very young children (aged 0-2) are more likely to use early childhood education and care services when they come from relatively advantaged socio-economic backgrounds; see OECD (2017).[1]

[1] In Ireland, the participation rate for children in low-income families is, at about 20%, less than one-third of that for children from high-income families (66%). In Belgium, France and the Netherlands, participation rates for children from low-income backgrounds are generally a little higher (around 30-40%), but are still only about half those for children from the richest families (roughly 60-75%). Similarly, in a number of OECD countries children are also more likely to use early childhood education and care when their mother is educated to degree-level. In the United Kingdom, the participation rate for children with a mother that has attained tertiary education is at 41%, 17 percentage points higher than the rate for children with mothers that have not attained tertiary education (24%). In Switzerland, the gap is as large as 30 percentage points.

Surprisingly, the interplay between child care provision/subsidization and redistribution has so far to a large extent been ignored in the literature.[2] We offer a fresh new look at this issue and propose a theoretical model whose crucial ingredient is an inefficient child penalty created by a gender norm. We then investigate the interaction between child penalties, child care policies and redistribution. Our research questions are the following. To what extent reducing the child care related gender inequalities and achieving a more equal income distribution are conflicting objectives? How can this potential conflict be mitigated by an appropriate design of the child care policies?

We consider a model in which spouses' career prospects are perfectly correlated. However, while fathers always enter a high-career path, mothers can either enter the same high-career path or a low-career one. In the latter case mothers are "traditional" because they are able to provide some informal child care. "Career mothers" instead need to purchase full-time formal care in the market. The sorting of women across career paths is endogenous and shaped by a social norm about gender roles in the family. Via this social norm traditional mothers' informal care imposes an externality on career mothers, who feel guilt if they provide less informal care than the average amount provided by woman. Hence, in the *laissez-faire* informal care is too large and the share of career mothers is too small. This translates in inefficiently high child penalties so that inefficiency and gender inequality go hand in hand. Furthermore, career choices exacerbate income inequalities (as measured for instance by the Gini coefficient) because higher incomes are concentrated on a smaller share of the population, which further decreases social welfare. We study the optimal design of linear and non-linear child care policies when the government is concerned with both child-care related efficiency and redistribution.

In a recent paper, Barigozzi *et al.* (2018) have examined the interplay between social norms, career choices and child-care decisions. We build on their model but adopt a different modeling strategy for the social norm. The research questions addressed in the two papers are completely different. Barigozzi *et al.* (2018) study whether eradicating or mitigating gender norms is socially optimal and how the design of specific policies (a uniform subsidy on child care, a women quota and parental leave) helps to achieve either one or the other objective. Redistribution is not a concern of the government in their model. In this paper we focus on the design of child care subsidies when income redistribution is relevant.

[2] Two exceptions are the literature on in-kind transfers and optimal taxation (Cremer and Pestieau 1996) and the literature on optimal taxation with endogenous fertility. In the latter, low-ability families may choose to 'specialize' in quantity, that is, to raise more children relative to higher-ability households. Child-related subsidies can, therefore, be used to enhance re-distribution: family size can be employed as an indicator for the earning capacity of the household (Cigno 1986). We totally depart from both strands of literature. We do not solve a model of optimal income taxation, we instead design non-linear child care subsidies. In addition, the number of children is exogenous in our model.

We show that, in a first-best, full information world efficiency and equity are separable. Redistribution is performed via lump-sum transfers and taxes which are designed to equalize utilities across all couples. Child care policies, on the other hand, are designed to achieve the appropriate level of informal child care and the efficient share of high career couples. Since the underlying problem is an externality, it is not surprising that the efficient policy involves a Pigouvian subsidy on market child care, which acts like a Pigouvian tax on informal care. And once child care levels are efficient, the induced career choices are also efficient. However, since this policy taxes away all extra earnings of high-career couples, it is of course not incentive compatible and it cannot be implemented when the spouses' earning opportunities in the high career path are not observable. This leads to the study of feasible second-best policies.

We consider two types of second-best settings. First, we study a linear subsidy and we show that it involves a trade-off between efficiency and redistributive considerations. Consequently, the optimal subsidy is lower than the Pigouvian level which applies when efficiency is the only social concerns.

More interestingly, we then show that this trade-off depends on the linearity of the policy. To see that we characterize the optimal incentive compatible policy, that is the non-linear policy constrained by the information structure. We show that this policy reestablishes the first-best "Pigouvian" rule for the (marginal) subsidy on informal care. In other words there is no longer a trade-off between child care subsidies and income redistribution. High-career couples enjoy positive rents and their share has to be reduced (compared to the first-best) to mitigate these rents. Consequently the outcome remains second-best. Still the policy is effective in reconciling the objectives of efficiency and income redistribution across couples. Note that the subsidy on formal care can be implicit in the case where child care is provided in kind.

The information requirement to implement this policy is rather minimal. It is sufficient that career paths *or* levels of formal child care are publicly observable. Amongst these the first one appears to be the least restrictive. When consumption of formal child care is observable for each couple, "topping-up" of child care provided in kind can be prevented, which in practice may appear difficult. But our analysis shows that when career paths are observable, topping up, is not a problem anyway. High career couples will then receive full time care (in kind or subject to a non linear subsidy) and they do not want to supplement this level by care paid at full market prices anyway. And due to the implicit or explicit subsidy, low career couples consume already more formal care then they would at market prices.

From a practical perspective, the non linearity or the policy introduces a measure of means-testing into our policies because child-care fees effectively differ across income levels. Because of the information limitations, means-testing remains quite basic and couples within a given career path cannot be distinguished. Still even this basic screening device has a rather dramatic impact in reconciling redistribution and child care policies (see also Sections 8 and 9 on this point). The child care subsidy is according to the first-best Pigouvian rule. Unlike in the linear case, there is no need to set it at a lower level for redistributive reasons.

In Section 7, we examine the robustness of our results with respect to the specification of the gender norm. We study the case where the norm is determined by the social group of the mother. In other words, social comparisons defining the norm are restricted to the specific group a mother belongs to, rather than being based on the entire population. For example, career mothers may feel guilt if they provide less informal care than women they interact with on a daily basis in their neighborhood. This definition of the norm appears to be more realistic. It also raises new interesting issues because now policies have to address income redistribution both across and within social groups. We show that our main message continues to hold: with a linear policy the tradeoff between efficiency and redistributive considerations persists whereas with a nonlinear policy the conflict basically disappears.

Finally, Section 8 examines how our results relate to real-world child policies and can provide guidance for policy design and reforms.

2 The model

Consider a population of couples with children, the size of which is normalized to one. Each couple consists of a mother 'm', a father 'f', and a given number of children. Couples choose their career path, the mode of child care, and their consumption.

There exist two types of career paths (indexed by j). First, a full engaging high-career path, $j = h$, where individuals who take up this career path have to work an entire day which we normalize to one. This constraint can be due to high peer pressure to work hard and to be fully committed, or to the obligation to spend the whole working day at the workplace—think for instance about a lawyer aiming to become a partner of the law firm.

Second, there is a less demanding low-career path, $j = \ell$, offering flexible working hours. Examples include most low qualified job but also some positions for college graduates like middle and high school teacher. The time not spent at work can be used for child care c_i, where $i = f, m$. Both jobs pay the wage rate y, but the high-career path comes with additional future earning possibilities q_i. We let $q_f \in [0, Q]$ and $q_m = \alpha q_f \in [0, \alpha Q]$, with $\alpha \in (0, 1]$. An $\alpha < 1$ captures pure discrimination: unequal pay for equally qualified workers, as it continues to be documented in nearly all developed countries.[3] Observe that while $\alpha < 1$ adds a measure of realism to the descriptive part of our model, it will not be essential for our results that all continue to hold

[3] The parameter α generates the unexplained component in the Oaxaca-Blinder decomposition of the GWG; see Blinder (1973) and Oaxaca (1973). Equation (4) below presents the decomposition of the GWG obtained in our model.

when $\alpha = 1$. Future revenue q_f is distributed according to the density function $f(.)$, with the cumulative distribution being $F(.)$. Future earning opportunities are perfectly correlated in a couple. Consequently, there is a single level of q_m associated with each level of q_f.[4]

Care for children provided by the spouse(s) is denoted by c_i $(i = f, m)$, while that bought in the private market is denoted by c_p. The latter costs p per unit of time. We let $p = y$, meaning that the current salary of one member in the couple exactly covers the costs of buying full-time child care on the private market.[5] The children must be taken care of for the entire day, implying $c_f + c_m + c_p = 1$. Couples in which both parents choose the high-career path thus have to fully rely on private child care. When parents enter a flexible job their salary decreases proportionally to the time devoted to care. Informal and private care constitute a family public good and its value to the parents is given by:

$$G(c_f, c_m, c_p) = v(c_f + c_m) + \beta v(c_p),$$

where $v' > 0, v'' < 0$ and $v(0) = 0$. Care provided by the father and mother are thus perfect substitutes while informal and private care are imperfect substitutes, with private care being (weakly) less welfare-enhancing than informal care, $\beta \in (0, 1]$.[6] Apart from child care, each parent derives utility from consumption of a numeraire commodity x.

Following Akerlof and Kranton (2000; 2010), individuals may suffer a disutility by deviating from the social categories that are associated with their identity (that is, an individual's sense of self), which causes behavior to conform toward those norms. We assume that mothers try to conform to the behavior of their peers. They feel guilt if they provide less informal care than the average level provided by women in the society.[7]

Given our assumption on the flexibility associated with the two available career paths, the social norm for mothers corresponds to the cost of the full-time job given by $\gamma(\max\{0; \bar{c}_m - c_m\})$, where $\bar{c}_m$ is the average time spent with children by mothers in the society. The parameter $\gamma \in [0, 1]$ reflects the costs of norm deviations. In Section 7 we extend the model to that case where social comparisons defining the norm are restricted to the specific group a mother belongs to rather then being based on the entire women population. Mothers then suffer a disutility by deviating from the "restricted" social categories that are "strongly" associated

[4] Assortative mating is commonly observed and has been increasing over the last decades; see Chiappori *et al.* (2017) and references within.

[5] This assumption is of no relevance to our results. Without it we would obtain a term proportional to $(p - y)$ in the first-order conditions with respect to child care. This would affect the equilibrium levels of child care but otherwise all other results are not affected.

[6] See, for instance, Gregg *et al.* (2005), Bernal (2008), and Huerta *et al.* (2011).

[7] The psychology literature points out that social norms on gender roles may cause mothers who work full-time to feel guilt when delegating the care of their children to others; see, Guendouzi (2006), Rotkirch and Janhunen (2010) and Rose (2017), among others.

with their identity. For example, they could refer to family members or to women they regularly interact with in their neighborhood. We show that our main results continue to hold under this alternative specification of the social norm.

The timing of couples' decisions is as follows: first, parents choose their career path and then, in the second stage, they choose consumption and the amount of formal and informal child care. Parents act cooperatively and maximize the sum of their utilities:

$$W = x_m + x_f + G(c_f, c_m, c_p) - \gamma(\max\{0; \bar{c}_m - c_m\}). \tag{1}$$

2.1 Couple's optimization

We first analyze the choice of child care activities for a given career path. Then, by proceeding backward, we consider the choice of career path made by the couple. This allows us to determine the average child care provided in the society and thus to define the cost of the social norm for mothers. We consider only decisions made at the second stage by the couples that turn out to be relevant for our analysis, namely the couples where *(i)* only the father enters the high-career path while the mother enters the flexible job market (traditional couples), and those where *(ii)* both parents take up the high-career path; see Appendix A.1 for the dominated couples' decisions.[8]

Traditional couple. Denote welfare of this couple by $W_{h\ell}$, where the first subscript refers to the father's career choice and the second subscript refers to the mother's career choice. Since the father took up the high-career path he is not able to take care of the children, and $c_f^* = 0$. Noting that $c_m + c_p = 1$, the couple chooses child care private provision to maximize (1) where $x_{h\ell} = x_m + x_f = y + q$ because $p = y$.[9] Optimal level of formal child care is thus implicitly determined by

$$\beta v'(c_p^*) = v'(1 - c_p^*). \tag{2}$$

First-order condition (2) indicates that traditional mothers purchase formal care, c_p^*, in the market up to the point where marginal utility from formal care equals the marginal benefit from informal care, $1 - c_p^*$.

The marginal norm cost for traditional mothers, γ, does not enter the FOC (2). To explain this, denote $c_{h\ell}^*$ and c_{hh} informal care provided by traditional parents and career parents, respectively. Traditional mothers do not suffer any norm cost because by definition we have $c_{hh} = 0$ so that $c_{h\ell}^* = 1 - c_p^* > \bar{c} > c_{hh}^* = 0$.

[8] Only the mother in the high-career path is dominated by having both parents entering the high-career path which involves higher future benefits. Similarly, having both parents entering the low-career path can never be optimal since then the couple forgoes future benefits q_f.

[9] Spouses' labor income, net of formal child care expenditures, is $x_{h\ell} = y + q + (1 - c_m)y - pc_p = y + q + c_p y - pc_p$.

The indirect utility of this $h\ell-$couple as a function of private child care c_p^* writes:

$$W_{h\ell}^* = y + q + v(1 - c_p^*) + \beta v(c_p^*)$$

High-career couple. High-career couples have no child care decision to make; they have to buy the full amount of private care on the market. High-career mothers suffer the cost from deviating from the norm and the couple's welfare amounts to:

$$W_{hh}^* = y + q(1 + \alpha) + \beta v(1) - \gamma \bar{c}.$$

Note that high-career couples who exclusively have to rely on private child care are those with higher consumption levels, that is $x_{h\ell}^* = y + q < x_{hh}^* = y + q(1 + \alpha)$.

We are now in the position to analyze the couple's decision about the two partners' career paths. Families have to choose whether to be a high-career $hh-$couple fully relying of formal child care, or to be a traditional $h\ell-$couple where the mother provides some informal care. A couple will become a high-career couple if it is beneficiary to do so, that is if $W_{hh}^* \geq W_{h\ell}^*$, or if

$$q \geq \hat{q}^* \equiv \frac{1}{\alpha}\left[v(1 - c_p^*) + \beta v(c_p^*) - \beta v(1) + \gamma \bar{c}\right].$$

The marginal couple $\hat{q}^*$ is the couple where parents are indifferent between belonging to a traditional and to a career couple. Given $\hat{q}^*$ we can now define average informal child care in society:

$$\bar{c} = \int_0^{\hat{q}^*} c_{h\ell}^* f(q) \mathrm{d}q = F(\hat{q}^*) c_{h\ell}^* = F(\hat{q}^*)(1 - c_p^*).$$

Note that by assuming a quasi-linear welfare function we consider the least favorable scenario for our argument. Indeed, considering a concave $u(x_m + x_f)$ would imply that formal care expenditures are increasing in q in traditional families. This would exacerbate the regressive effect of child care subsidies.[10]

2.2 Market outcome

An allocation is given by the identity of the marginal couple and by the amount of child care provided by traditional couples. The following proposition characterizes the *laissez-faire* allocation.

[10] To see why one can derive $\frac{dc_p}{dq} = \frac{pu''(x_m + x_f)}{SOC}$, where:

$$SOC = pu''(2y + q - pc_p) + v''(1 - c_p) + \beta v''(c_p) < 0.$$

Hence formal child care is strictly increasing in household income for $q \leq \hat{q}^*$ and is weakly increasing in household income for $q > \hat{q}^*$ because, by assumption, $c_p = 1$ for high career couples.

Proposition 1 (Characterization of the *laissez-faire*) *When mothers who do not provide child care suffer from deviating from the social norm, i.e.* $\gamma > 0$*, and/or the job market suffers from gender discrimination,* $\alpha < 1$*, then:*

(i) it is never optimal for fathers to take up the low-career path;

(ii) the marginal couple is given by

$$\hat{q}^* = \frac{1}{\alpha}\left[v(1-c_p^*) + \beta\left[v(c_p^*) - v(1)\right] + \gamma F(\hat{q}^*)(1-c_p^*)\right], \tag{3}$$

couples with future job opportunities higher or equal to the threshold $\hat{q}^*$ *choose the high-career path for both parents;*

(iii) private care purchased by traditional couples, c_p^**, satisfies equation (2).*

There are both traditional and career couples in the economy if $\hat{q}^* \in (0, Q)$. This is the most interesting case as it implies that the social norm persists as it is the case in most current societies, albeit to a different extent in different countries. We concentrate on this case, even though other outcomes are possible, depending on the parameters of the model and the distribution $F(q)$.[11]

From (3), an interior solution requires that $\hat{q}^*$ exists such that $\hat{q}^* = (1/\alpha)[v(1-c_p^*) + \beta\left[v(c_p^*) - v(1)\right] + \gamma F(\hat{q}^*)(1-c_p^*)] < Q$. Due to the concavity of $v(\cdot)$, $v(1-c_p^*) + \beta\left[v(c_p^*) - v(1)\right] > 0$ holds so that the previous inequality is always met provided that Q is sufficiently large and $F(\hat{q})$ is concave, which we assume in the remainder of the paper. This also ensures that the marginal couple $\hat{q}^*$ is unique.

The gender wage gap (GWG) is defined as the *difference in total income earned by mothers and fathers* in equilibrium and is given by:

$$\begin{aligned} GWG &= \int_0^Q [y+q]f(q)\mathrm{d}q - \left[F(\hat{q}^*)yc_p^* + \int_{\hat{q}^*}^Q [y+\alpha q]f(q)\mathrm{d}q\right] \\ &= \underbrace{F(\hat{q}^*)\left(1-c_p^*\right)y}_{\text{child penalty}} + \underbrace{\int_0^{\hat{q}^*} qf(q)\mathrm{d}q}_{\text{adverse sorting}} + \underbrace{\int_{\hat{q}^*}^Q (1-\alpha)qf(q)\mathrm{d}q}_{\text{plain discrimination}} \end{aligned} \tag{4}$$

The GWG decomposes in the gap between the hours worked because of family duties, and in the different return to labor supplied in sectors where man and women are employed. The first term in (4) thus represents "child penalty" (see Blau and Kahn 2017; Kleven *et al.* 2018): mothers in traditional couples do not work full time, but spend part of their time to provide

[11] If instead $\hat{q}^* \geq Q$ no one is suffering the utility loss because all couples are traditional. A situation such that $\hat{q}^* \leq 0$ corresponds to the case where the social norm does not play any role because only career couples exist. This maximizes labor income but also implies a welfare loss because of the forgone utility coming from informal care. Though theoretically possible, none of these cases appears to be empirically relevant.

informal child care. Child penalty thus depends on average informal care, $\bar{c} = F(\hat{q}^*)\left(1 - c_p^*\right)$, provided by traditional mothers. The second term accounts for the fact that women forego the extra earning opportunities associated with the high-career path. Interestingly, both child penalty and "adverse sorting" are affected by social norms and child care decision through $\hat{q}^*$. They decrease when the share of career mothers in the society increases. The model thus offers a clean explanation of how social pressure determines women sorting and thus their low participation in leading positions together with lower wages. Finally, the last term in (4) captures the unexplained component of the GWG of the Oaxaca–Blinder decomposition, or the plain discrimination part; it vanishes when $\alpha = 1$.

Before turning to the design of child-care policy, we define the social planner's objective function and the optimal allocation.

3 The optimal allocation

The social planner is interested both in efficiency and in redistribution. Specifically, the social welfare function is assumed to be a concave transformation, $\Psi(\cdot)$, of the families' welfare functions in order to capture inter-family inequality aversion.[12] Thus, a first-best (fb) allocation is defined by aggregate consumption levels $x_{h\ell}^{fb}(q)$ and $x_{hh}^{fb}(q)$, by the indifferent couple, $\hat{q}^{fb}$ (which determines the share of female participation in the high-career path), and by the level of formal child care chosen by traditional couples, $c_p^{fb}(q)$ for $q < \hat{q}^{fb}$ (recall that, by definition, $c_p^{fb}(q) = 1$ for $q \geq \hat{q}^{fb}$).

Specifically, the social planner chooses $\{x_{hh}(q), x_{h\ell}(q), c_p(q), \hat{q}\}$ to maximize the following welfare function:

$$SW = \int_0^{\hat{q}} \Psi\big(x_{h\ell}(q) + v(1 - c_p(q)) + \beta v(c_p(q))\big) f(q) \mathrm{d}q + \int_{\hat{q}}^{Q} \Psi\big(x_{hh}(q) + \beta v(1) - \gamma \bar{c}\big) f(q) \mathrm{d}q \tag{5}$$

subject to the budget constraint:

$$y + \int_0^Q q f(q) \mathrm{d}q + \int_{\hat{q}}^Q \alpha q f(q) \mathrm{d}q = \int_0^{\hat{q}} x_{h\ell}(q) f(q) \mathrm{d}q + \int_{\hat{q}}^Q x_{hh}(q) f(q) \mathrm{d}q, \tag{6}$$

where $\bar{c} = \int_0^{\hat{q}} (1 - c_p(q))\, f(q)\, \mathrm{d}q$.

In Appendix A.2 we derive the optimal allocation that is characterized as follows. Welfare is constant irrespective of the couple's career path and their future earning possibilities:

$$W_{h\ell}^{fb}(q) = W_{hh}^{fb}(q) = W^{fb} \quad \forall q;$$

[12] In Barigozzi *et al.* (2018), redistribution across income levels is not relevant because they assume quasi-linear preferences with a constant marginal utility of income. While the excessive share of traditional couples does also affect the income distribution by making it more concentrated this in itself does not affect welfare in their setting.

Formal child care is such that $c_p^{fb}(q) = c_p^{fb}\ \forall q$ and is implicitly given by:

$$\beta v'(c_p^{fb}) + [1 - F(\hat{q}^{fb})]\gamma = v'(1 - c_p^{fb}). \tag{7}$$

The left-hand side denotes the social marginal benefit of formal child care while the right-hand side denotes the social marginal cost of informal care. Note that the above equation is independent of a traditional couple's q. Compared to the *laissez-faire* described in (2), the marginal benefit contains an additional term $[1-F(\hat{q}^{fb})]\gamma$ which reflects the negative externality of informal care provision on type-hh couples whose share is $1 - F(\hat{q}^{fb})$. Informal child care is thus inefficiently high in the *laissez-faire*, which translates in underconsumption of formal care: $c_p^* < c_p^{fb}$. Not surprisingly c_p^{fb} and $\hat{q}^{fb}$ do not depend on the social welfare function Ψ. This is due to the quasi-linearity of preferences. All Pareto-efficient allocations imply the same levels of c_p and $\hat{q}$, but *may* differ in consumption levels. But since we use a symmetric social welfare function any concave Ψ implies that in the first-best utility levels are equalized. However, the degree of concavity will matter in the second-best settings considered below.

Interestingly, $W_{h\ell}^{fb}(q) = W_{hh}^{fb}(q)$ and $c_p^{fb}(q) = c_p^{fb}\ \forall q$ imply that the consumption of couples is constant in each career-path: $x_{h\ell}^{fb}(q) = x_{h\ell}^{fb}$ and $x_{hh}^{fb}(q) = x_{hh}^{fb}\ \forall q$. This in turn implies that:

$$x_{hh}^{fb} - x_{h\ell}^{fb} = v(1 - c_p^{fb}) + \beta[v(c_p^{fb}) - v(1)] + \gamma F(\hat{q}^{fb})(1 - c_p^{fb}) > 0 \tag{8}$$

The above expression shows that high-career couples do not get higher consumption because of their higher q (as it was the case in the *laissez-faire*), but because the government compensates them for their utility loss due to the cost of the social norm and to the purchase of full private care (whose utility is mitigated by the parameter β). Finally, in Appendix A.2 we show that the FOC wrt $\hat{q}$ can be rewritten as:

$$\alpha\hat{q}^{fb}f(\hat{q}^{fb}) = f(\hat{q}^{fb})[v(1-c_p^{fb})+\beta(v(c_p^{fb})-v(1))+\gamma F(\hat{q}^{fb})(1-c_p^{fb})]-\gamma[1-F(\hat{q}^{fb})](1-c_p^{fb})f(\hat{q}^{fb}) \tag{9}$$

so that

$$\hat{q}^{fb} \equiv \frac{1}{\alpha}\{[v(1 - c_p^{fb}) + \beta(v(c_p^{fb}) - v(1))] + \gamma F(\hat{q}^{fb})(1 - c_p^{fb}) - \gamma[1 - F(\hat{q}^{fb})](1 - c_p^{fb})\} \tag{10}$$

Comparing (3) and (10) and recalling that $c_p^* < c_p^{fb}$, we observe that $\hat{q}^* > \hat{q}^{fb}$, that is the share of high-career couples is inefficiently low in the *laissez-faire*.

Expression (9) has a simple interpretation in terms of cost and benefits of *decreasing* $\hat{q}$ (that is moving $f(\hat{q})$ couples from traditional to high-career). The LHS measures the marginal benefits in terms of extra future earnings. In the RHS, the first two terms in brackets represent the net lost utility from formal care and the norm cost, respectively. The last term is the Pigouvian term which is negative because the externality imposed on all high-career couples decreases because the average informal care falls. Formally, we have $\partial\bar{c}/\partial\hat{q} = (1 - c_p)f(\hat{q})$. Since a negative cost is effectively a benefit this term could have been moved to the LHS, but since the interpretation of (9) also shows that of (10) this presentation is more telling.[13]

Observe that $\hat{q}^{fb}$ does not depend on Ψ; it is the same in *all* Pareto efficient allocations. The first-best level $\hat{q}^{fb}$ is set purely on efficiency grounds—to maximize the size of the cake which is then redistributed according to social preferences (which in our case involves equalization of utilities).

The following propositions characterizes the optimal allocation:

Proposition 2 (The optimal allocation) *The optimal allocation* $\{x_{hh}^{fb}, x_{h\ell}^{fb}, c_p^{fb}, \hat{q}^{fb}\}$ *maximizes the social welfare function (5) subject to the budget constraint (6) and is characterized as follows:*

(i) Couples' welfare does not depend on the career choice of the mother nor on career prospects: $W_{h\ell}^{fb}(q) = W_{hh}^{fb}(q)\ \forall\ q$. *High-career couples get higher consumption because they are compensated for their utility loss due to full private care and due to the cost of the social norm.*

(ii) Formal child care $c_p^{fb}(q) = c_p^{fb}$ *is the same for all traditional couples and satisfies (7). It is chosen such that the negative externality induced by the social norm is fully internalized.*

(iii) The share of high-career couples is given by $1 - F(\hat{q}^{fb})$ *where the marginal couple* $\hat{q}^{fb}$ *is defined in (10).*

(iv) The optimal level of the GWG entails a child penalty and a sorting differential equivalent to $F(\hat{q}^{fb})\left(1 - c_p^{fb}\right) y$ *and* $\int_0^{\hat{q}^{fb}} qf(q)dq$*, respectively.*

Point (iv) directly follows from substituting $(c_p^{fb}, \hat{q}^{fb})$ into equation (4).

3.1 Welfare analysis of the *laissez-faire* allocation

By comparing the optimal allocation and the market outcome we can establish in which sense the *laissez-faire* allocation is inefficient.

Proposition 3 (Welfare analysis of the *laissez-faire*) *In the laissez-faire allocation:*

(i) Within each career path, welfare differs across couples; it increases with career prospect q.

(ii) Formal child care, c_p^**, is inefficiently low and informal care,* $c_{h\ell}^*$*, is too high. This is due to the negative externality that informal care exerts on high-career mothers through the social norm.*

[13] Similarly, multiplying both sides of (9) by -1, would be more in line with the original FOC, because it then measures the cost and benefits (reversed from the interpretation discussed) of increasing $\widehat{q}$.

(iii) Female participation in the high-career path is inefficiently low, $\hat{q}^{fb} < \hat{q}^*$.

(iv) In the GWG, both the child penalty and adverse sorting are inefficiently high.

In the first-best all couples receive the same welfare. Proposition 3*(i)* shows that, in the laissez-faire, welfare is increasing in q both among traditional couples and among career couples. Thus, welfare is equalized neither across couples belonging to different career paths nor across couples within the same career path.

Point *(ii)* shows that the negative externality translates into under consumption of formal child care by traditional couples in *laissez-faire* ($c_p^* < c_p^{fb}$). Point *(iii)* concerns the share of women entering the high-career path which is inefficiently low in *laissez-faire*. When the negative externality is internalized, formal child care increases and the cost of the social norm falls. As a result the high-career path is more attractive in the first-best, or $\hat{q}^* > \hat{q}^{fb}$.

Finally, point *(iv)* requires some explanations. For any given q, in the *laissez faire*, the female spouse's earnings are less than or equal to her first-best earnings. Indeed, child penalty is lower in the first-best because women's labor income is higher due to the higher formal child care ($c_p^* < c_p^{fb}$). The optimal level of child penalty is thus obtained when the negative externality exerted by traditional mothers on career mothers is properly taken into account. This clarifies why, in the model, efficiency is reached via the appropriate reduction of child care related inequalities. Finally, adverse sorting is lower because more women enter the high-career path and benefit from future prospects ($\hat{q}^* > \hat{q}^{fb}$).

4 Decentralizing the first-best allocation

Decentralization of the first-best solution requires a subsidy s on formal child care and individualized lump-sum taxes or transfers $T_{h\ell}(q)$ and $T_{hh}(q)$. When a subsidy s is in place, the net price of private child care is $p^n = p - s = y - s$, and a traditional couple's optimal child care decision solves:

$$v'(1 - c_p) - s = \beta v'(c_p). \tag{11}$$

Comparing (11) with (7) shows that a subsidy of

$$s^{fb} = [1 - F(\hat{q}^{fb})]\gamma \tag{12}$$

implements the first-best level of child care. Since formal and informal care sum up to one, a subsidy on market care is effectively a tax on informal care. According to equation (12) s^{fb} corresponds to a Pigouvian *tax* on informal child care; it equals the marginal social cost of the externality informal care imposes on high-career couples.

The lump-sum transfers $T_{h\ell}(q)$ and $T_{hh}(q)$ must be chosen such that welfare levels between all couples are equalized, that is

$$W_{h\ell}(q) = y + q + s^{fb}c_p^{fb} + v(1 - c_p^{fb}) + \beta v(c_p^{fb}) + T_{h\ell}(q) = W^{fb} \quad \text{when} \quad q \leq \hat{q}^{fb},$$
$$W_{hh}(q) = y + (1+\alpha)q + s^{fb} + \beta v(1) - \gamma\bar{c} + T_{hh}(q) = W^{fb} \quad \text{when} \quad q \geq \hat{q}^{fb}.$$

Decentralizing $\hat{q}^{fb}$ further requires $T_{h\ell}(\hat{q}^{fb}) = T_{hh}(\hat{q}^{fb})$. To see this note that when $T_{h\ell}(\hat{q}) = T_{hh}(\hat{q})$ the marginal couple defined by $W_{h\ell}(\hat{q}) = W_{hh}(\hat{q})$ is determined by

$$\begin{aligned} & y + \hat{q} + s^{fb}c_p^{fb} + v(1 - c_p^{fb}) + \beta v(c_p^{fb}) \\ & = y + (1+\alpha)\hat{q} + s^{fb} + \beta v(1) - \gamma\bar{c} \\ \Leftrightarrow \quad & \hat{q} = \frac{1}{\alpha}[v(1 - c_p^{fb}) + \beta\left(v(c_p^{fb}) - v(1)\right) + \gamma\bar{c} - s^{fb}(1 - c_p^{fb})] \end{aligned} \tag{13}$$

Using (12) together with $\gamma\bar{c} = \gamma F(\hat{q})(1 - c_p^{fb})$ shows that (13) and (10) coincide once formal child care is subsidized at the Pigouvian rate.

Hence, with sufficiently powerful instruments efficiency and redistribution can be addressed separately: the Pigouvian subsidy s^{fb} on private child care optimally reduces informal child care provision (and thus child penalties) while the transfers $T_{h\ell}(q)$ and $T_{hh}(q)$ assure equal welfare to all couples. Note that the individualized transfers redistribute from high to low q couples but also compensate the high-career couples for their utility losses due to full private care and to their cost of the social norm.[14]

We now turn to the study of second-best policies.

5 Linear policy

First, we consider a simple policy under which instruments are restricted in an *ad hoc* way. In other words, we remain agnostic about the information structure. We assume that the instruments necessary to implement the first-best are not available (specifically the individualized transfers) and consider a simple policy which is empirically appealing and effectively used in practice.

The considered policy consists of a uniform (linear) subsidy s on market child care, financed by a uniform lump-sum tax τ. The government's budget constraint is then given by

$$\tau = sF(\hat{q}(p^n))c_p(p^n) + s[1 - F(\hat{q}(p^n))]. \tag{14}$$

[14] In Barigozzi *et al.* (2018), the social norm is determined by child-care decisions made by the median couple of the preceding generation. With this different specification of the norm it turns out that a Pigouvian subsidy does not restore efficiency but reduces informal care too much. Hence the optimal subsidy must be set below the Pigouvian rule.

Recall that $p^n = p - s = y - s$ is the net, after subsidy, price of market care. Let us denote $c_p^s = c_p(p^n)$ consumption of formal care under the linear subsidy s. As before it is implicitly determined by:

$$v'(1 - c_p^s) - s = \beta v'(c_p^s) \tag{15}$$

The social welfare function can be written as:

$$\begin{aligned} SW(s,\tau) = &\int_0^{\hat{q}(p^n)} \Psi\big(y + q + sc_p(p^n) - \tau + v(1 - c_p(p^n)) + \beta v(c_p(p^n))\big) f(q)\mathrm{d}q \\ &+ \int_{\hat{q}(p^n)}^{Q} \Psi\big(y + (1+\alpha)q + s - \tau + \beta v(1) - \gamma \bar{c}(p^n)\big) f(q)\mathrm{d}q, \end{aligned} \tag{16}$$

where $\hat{q}^s = \hat{q}(p^n)$ and $\bar{c}(p^n) = F(\hat{q}^s)(1 - c_p(p^n))$. The FOC wrt τ is given by:

$$\lambda = \int_0^Q \Psi'(q) f(q)\mathrm{d}q \equiv E[\Psi'], \tag{17}$$

where λ is the Lagrangean multiplier of the budget constraint, E is the expectation operator and $\Psi(q)$ is defined as $\Psi(W_{h\ell}(q))$ for $h\ell$ couples and as $\Psi(W_{hh}(q))$ for hh couples. This equation has a familiar flavor from linear taxation models, in particular Sheshinski (1972). It states that the social marginal cost of raising an additional dollar, λ, should be equal to its social marginal benefit, $E[\Psi']$. Now define:

$$E_{h\ell}[\Psi'] \equiv \frac{\int_0^{\hat{q}^s} \Psi'(q) f(q)\mathrm{d}q}{F(\hat{q}^*)} \quad \text{and} \quad E_{hh}[\Psi'] \equiv \frac{\int_{\hat{q}^s}^{Q} \Psi'(q) f(q)\mathrm{d}q}{1 - F(\hat{q}^*)}, \tag{18}$$

which represent the average marginal utilities of income by traditional and high-career couples respectively.

Considering that $\partial p_n / \partial s = -1$, the FOC with respect to s can be written as:

$$\begin{aligned} &F(\hat{q}^s)E_{h\ell}[\Psi']c_p^s + (1 - F(\hat{q}^s))E_{hh}[\Psi']\left[1 - \gamma F(\hat{q}^s)\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} + \gamma(1 - c_p^s) f(\hat{q}^s)\frac{\mathrm{d}\hat{q}^s}{\mathrm{d}p^n}\right] \\ &- \lambda\left[F(\hat{q}^s)c_p^s - sF(\hat{q}^s)\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} + s(1 - c_p^s) f(\hat{q}^s)\frac{\mathrm{d}\hat{q}^s}{\mathrm{d}p^n} + 1 - F(\hat{q}^s)\right] = 0. \end{aligned} \tag{19}$$

Noting that $E[c_p^s] = F(\hat{q}^s)c_p^s + 1 - F(\hat{q}^s)$ we show in Appendix A.4 that the optimal linear subsidy on formal child care, s^o, amounts to:

$$s^o = \gamma \frac{(1 - F(\hat{q}^s))E_{hh}[\Psi']}{E[\Psi']} - \frac{\mathrm{cov}[\Psi', c_p^s]}{E[\Psi']\frac{\partial E[c_p^s]}{\partial p^n}} \tag{20}$$

The first expression is the Pigouvian term and the second term is the redistributive term. When $\Psi'' = 0$ so that social welfare is not concave and there is no concern for redistribution and the above expression reduces to $s^o = [1 - F(\hat{q}^s)]\gamma$, which is the first-best Pigouvian rule. From expression (13) this also yields $\hat{q} = \hat{q}^s$ so that we return to the first-best allocation. When the social welfare function is concave, we have $\text{cov}[\Psi', c_p^s] < 0$ since families with higher formal care have a higher welfare. In the Appendix we show that $\partial E[c_p^s]/\partial p^n < 0$ so that the second term on the RHS in expression (20) is negative (a positive fraction is preceded by a negative sign). Redistributive concerns thus decrease optimal child care subsidies since it is mainly the high-career couples who profit from such subsidies. Furthermore, we have $E_{hh}[\Psi'] < E[\Psi']$ so that the Pigouvian term is also reduced compared to its first-best counterpart. This is because the externality affects high career-couples who in the second-best have a lower social marginal utility. The marginal social damage of the externality is determined by converting their (marginal) utility into social (marginal) utility, which is achieved by the term $E_{hh}[\Psi']/E[\Psi']$.[15] Consequently, we have $s^o < s^{fb}$; see Appendix A.4 for the formal proof.

Proposition 4 (Linear child care subsidy) *The optimal linear policy when redistribution is relevant ($\Psi'' > 0$) implies:*

(i) $s^o < s^{fb}$ *because it is mainly the high-career couples who profit from this policy. Thus, formal child care purchased by traditional couples,* c_p^s*, is inefficiently low* $\left(c_p^{fb} > c_p^s\right)$*;*

(ii) and $\hat{q}^s > \hat{q}^{fb}$ *so that there are more traditional couples in the second best than in the first-best. The marginal couple is distorted upwards to reduce the share of high career couples receiving the subsidy for full-time formal care which improves redistribution.*

(iii) In the GWG, both child penalties and adverse sorting are inefficiently high.

The intuition for *(iii)* is the same as for the corresponding point in Proposition 3. As expected, the linear subsidy mitigates the inefficiency of the *laissez-faire* informal care provision but does not fully restore efficiency. However, welfare is obviously higher with the linear policy than in the *laissez faire.*

6 Nonlinear policy

Now, we take a different approach and assume that the available policies are not restricted in an *ad hoc* way. Instead, we study the design of the best policy that is available *given the information structure.* This is not just a matter of theoretical interest. The important underlying practical question is whether the distortions characterized in the previous section are unavoidable once redistribution under asymmetric information is involved, or whether they are simply artifacts of the linearity of the considered policy.

Under full information this approach yields the first-best, but this supposes that all relevant variables, including a couple's high-career earning opportunities q are publicly observable. We

[15] In the FB, utilities are equalized so that this term is equal to one.

shall now assume that q is not publicly observable but that both the career path and the level of market care *are* observable at the individual (couple's) level. The government can then offer two contracts conditioned on the *reported* type $\tilde{q}$ denoted by $\{J(\tilde{q}), c_p^g(\tilde{q}), T(\tilde{q})\}$, where $J \in \{h\ell, hh\}$ indicates the career path, T is the transfer that households have to pay and $c_p^g(\tilde{q})$ is the amount of formal child care provided by the government. Since $c_p^g(\tilde{q})$ is observable at the couple's level, the distinction between in-kind provision and a nonlinear taxation of market care is not relevant; see Cremer and Gahvari (1997). To be more precise, this is simply a matter of practical implementation of the underlying optimal contract. This implies, in particular, that when $c_p^g(\tilde{q})$ is interpreted as in-kind provision, topping up is not possible.[16] As usual we shall, without loss of generality, concentrate on incentive compatible contracts.

Given that no topping up is possible it must be $c_p^g(q) = 1$ for all hh couples. In addition, given that, conditional on the career path, all families have the same preferences for child care, it is impossible to separate families according to q once the career path has been assigned. Hence, the government offers only two contracts: $\{T_{h\ell}, c_p^g\}$ for $h\ell$-couples and $\{T_{hh}, 1\}$ for hh-couples. In other words, all traditional couples consume the same level of market care and face the same tax or transfer. The same is true for all high-career couples.[17]

The average informal care provided by traditional mothers now is $\bar{c} = F(\hat{q}^g)(1 - c_p^g)$, where $\hat{q}^g$ indicates future prospects of the marginal couple, or the couple such that welfare is the same in the two career paths, $\hat{q}^g : W_{hh}(\hat{q}^g) = W_{h\ell}(\hat{q}^g)$.

The government maximizes the following welfare function:

$$\max_{T_{h\ell}, c_p^g, T_{hh}, \hat{q}^g} SW = \int_0^{\hat{q}^g} \Psi(\underbrace{y + q + c_p^g y - T_{h\ell} + v\left(1 - c_p^g\right) + \beta v\left(c_p^g\right)}_{W_{h\ell}}) f(q)\,\mathrm{d}q$$

$$+ \int_{\hat{q}^g}^{Q} \Psi(\underbrace{2y + (1+\alpha) q - T_{hh} + \beta v(1) - \gamma F(\hat{q}^g)\left(1 - c_p^g\right)}_{W_{hh}}) f(q)\,\mathrm{d}q \tag{21}$$

subject to the budget constraint

$$F(\hat{q}^g) T_{h\ell} + [1 - F(\hat{q}^g)] T_{hh} - p\left[F(\hat{q}^g) c_p^g + 1 - F(\hat{q}^g)\right] \geq 0, \tag{22}$$

and subject to the following incentive constraint:

$$\begin{aligned} & 2y + (1+\alpha)\hat{q}^g - T_{hh} + \beta v(1) - \gamma F(\hat{q}^g)\left(1 - c_p^g\right) \\ & - \left(y + \hat{q}^g + c_p^g y - T_{h\ell} + v\left(1 - c_p^g\right) + \beta v\left(c_p^g\right)\right) = 0. \end{aligned} \tag{23}$$

[16] With the considered information structure it can be prevented and nothing can be gained by allowing it.

[17] This is a well known property in contract theory and we skip the proof. To establish the results formally one has to maximize social welfare subject to the budget and incentive constraints. A simple first-order approach will show that the solution involves pooling within each career group.

Since there is pooling in both groups, incentive compatibility requires simply that $\hat{q}^g$ is indifferent between the two career paths. This follows because $\partial W_{hh}(q)/\partial q = 1 + \alpha > \partial W_{h\ell}(q)/\partial q = 1$ so that W_{hh} increases faster in q than $W_{h\ell}$. Consequently, condition (23) ensures that no high-career couple with future earnings $q \geq \hat{q}^g$ should have an incentive to mimic a traditional couple, that is $W_{hh}(q) \geq W_{h\ell}(q)\ \forall\ q \in [\hat{q}^g, Q]$. Similarly, it implies that no traditional couple wants to mimic a high career couple.

We denote the Lagrangian multipliers associated with the budget constraint and the incentive constraint $\widehat{\lambda}$ and μ respectively. Using the expectation operators defined in (18) we can write the FOCs with respect to the transfers $T_{h\ell}$ and T_{hh} as:

$$- E_{h\ell}[\Psi^{'}]F\left(\hat{q}^g\right) + \mu + \widehat{\lambda}F(\hat{q}^g) = 0 \tag{24}$$

$$- E_{hh}[\Psi'] \left[1 - F\left(\hat{q}^g\right)\right] - \mu + \widehat{\lambda}\left[1 - F\left(\hat{q}^g\right)\right] = 0 \tag{25}$$

Combining (24) and (25) and rearranging yields:

$$\widehat{\lambda} = \int_0^{\hat{q}^g} \Psi'\left(\cdot\right) f(q) \mathrm{d}q + \int_{\hat{q}^g}^{Q} \Psi'\left(\cdot\right) f(q) \mathrm{d}q = E\left[\Psi'\right]. \tag{26}$$

This equation simply states that the marginal cost of raising additional revenue, $\widehat{\lambda}$, must be equal to its marginal social benefit, $E[\Psi']$. The FOC with respect to formal child care for traditional couples, c_p^g, is given by:

$$\int_0^{\hat{q}^g} \Psi'\left(\cdot\right) \left[y - v'\left(1 - c_p^g\right) + \beta v'\left(c_p^g\right)\right] f\left(q\right) \mathrm{d}q + \int_{\hat{q}^g}^{Q} \Psi'\left(\cdot\right) \gamma F\left(\hat{q}^g\right) f(q) \mathrm{d}q$$

$$- \widehat{\lambda} p F\left(\hat{q}^g\right) + \mu\left[\gamma F\left(\hat{q}^g\right) - y + v'\left(1 - c_p^g\right) - \beta v'\left(c_p^g\right)\right] = 0 \tag{27}$$

In Appendix A.5 we show that by using (25) and (26) the (27) reduces to:

$$v'(1 - c_p^g) + \beta v(c_p^g) = [1 - F(\hat{q}^g)]\gamma. \tag{28}$$

Comparing this expression to (11) shows that the level of child care c_p^g can be decentralized by a subsidy on market care given by:

$$s^g = [1 - F(\hat{q}^g)]\gamma. \tag{29}$$

Consequently, the public provision of c_p^g corresponds to an implicit subsidy on market care which is set according to the Pigouvian *rule* defined by (12). In other words, it reflects the marginal social damage which is here measured by the extra norm cost imposed on all career couples. This is an interesting result because it implies that the downward distortion on s implied by the redistributive bias obtained in the previous section indeed appears to be an artifact of the *ad hoc* restrictions imposed on the policy, namely its simple linear specification. When the policy is constrained only by the information structure this distortion vanishes. However, while s^g is set according to the first-best Pigouvian *rule*, its actual *level* will differ from s^{fb}, unless $\hat{q}^g = \hat{q}^{fb}$. This brings us to the next question namely the comparison between $\hat{q}^g$ and $\hat{q}^{fb}$. This amounts to studying whether the solution under asymmetric information involves a distortion on the marginal couple and if yes in which direction.

The FOC with respect to $\hat{q}^g$ can be written as:

$$\Psi(W_{h\ell}(\hat{q}^g)) - \Psi(W_{hh}(\hat{q}^g)) - \gamma f(\hat{q}^g)(1-c_p^g)(1-F(\hat{q}^g))E_{hh}[\Psi']$$
$$+\widehat{\lambda}\left[f(\hat{q}^g)(T_{h\ell}-T_{hh}) + pf(\hat{q}^g)\left(1-c_p^g\right)\right] + \mu\left[\alpha - \gamma f(\hat{q}^g)\left(1-c_p^g\right)\right] = 0, \tag{30}$$

where the first two terms vanish because of the incentive constraint.

The approach is to evaluate the FOC for $\hat{q}^g$ at $\hat{q}^{fb}$ while adjusting all the other endogenous variables according to their respective FOCs.[18] When $\hat{q}^g = \hat{q}^{fb}$ we have from (28) that $c_p^g = c_p^{fb}$. In Appendix A.6 we show that

$$T_{h\ell} - T_{hh} = \gamma[1-F(\hat{q}^{fb})](1-c_p^g) - y(1-c_p^g). \tag{31}$$

Solving (25) for μ and inserting (31) in (30), we have:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \hat{q}^g}\Big|_{\hat{q}^g=\hat{q}^{fb}} =& - E_{hh}[\Psi'](1-F(\hat{q}^g))\gamma f(\hat{q}^{fb})(1-c_p^g)\\ &+ E[\Psi']\left[f(\hat{q}^{fb})(-y(1-c_p^g) + \gamma(1-F(\hat{q}^{fb}))(1-c_p^g)) + f(\hat{q}^{fb})y(1-c_p^g)\right]\\ &+ [-E_{hh}[\Psi'](1-F(\hat{q}^g)) + E[\Psi'](1-F(\hat{q}^{fb}))][\alpha - \gamma f(\hat{q}^{fb})(1-c_p^g)]\\ =&(1-F(\hat{q}^{fb}))(E[\Psi'] - E_{hh}[\Psi'])\alpha > 0.\end{aligned} \tag{32}$$

So that we have $\hat{q}^g > \hat{q}^{fb}$. In words, the second-best solution implies an upward distortion of the marginal couple $\hat{q}^g$. Consequently, there are more traditional couples in the second-best solution than in the first-best.

To understand this expression note that a couple with $q \geq \widehat{q}$ enjoys an informational rent of $\alpha(q-\widehat{q}) = W_{hh}(q) - W_{hh}(\widehat{q})$. Total rents are thus given by:

$$R = \int_{\widehat{q}}^{Q} \alpha(q-\widehat{q})f(q)\mathrm{d}q$$

and we have:[19]

$$\frac{\partial R}{\partial \widehat{q}} = -\alpha\int_{\widehat{q}}^{Q} f(q)\mathrm{d}q = -\alpha[1-F(\widehat{q})].$$

[18] If the other variables were held constant the sign of the derivative would be inconclusive. However, adjusting all the other variables in an optimal way reduces the problem to a single dimension so that the derivative is informative. As an example, consider the maximization of $f(x,y)$ and denote the solution (x^*, y^*). Showing that at any given point (x,y), $\partial f/\partial x > 0$ is not enough to show that $x > x^*$. However, by using the FOC for y we reduce the problem to the maximization of $f(x, y^*(x))$ and the derivative of this expression allows us to compare x and x^*, as long as the problem is concave which we have to assume anyway.

[19] Note that the derivative wrt the lower bound is zero.

Under full information these rents can be extracted and redistributed. Under asymmetric information they cannot because of the incentive constraint. As $\widehat{q}$ increases the extra amount $\alpha[1 - F(\widehat{q})]$ can be extracted and redistributed which implies a social benefit of $(E[\Psi'] - E_{hh}[\Psi'])\alpha(1 - F(\hat{q}^{fb}))$. In words, the second-best solution involves an upward distortion in the marginal couple in order to reduce "informational rents" of the high-career couples. This means that by increasing the level of q of the marginal couple more tax revenue can be extracted from the high-career couple and redistributed to the traditional couples with lower income, so that welfare increases.

We can now also return to the *levels* of the implicit subsidy implied by the policy. Equation (12) and (29) together with $\hat{q}^g > \hat{q}^{fb}$ imply $s^g < s^{fb}$, so that asymmetric information leads to a lower implicit subsidy on formal care. Intuitively, the strict Pigouvian rule applies in both cases but with $\hat{q}^g > \hat{q}^{fb}$ the group of high-career couples affected by the externality is smaller so that its marginal social damage is also smaller. Consequently, using (11) we'll also have $c_p^g < c_p^{fb}$. As in the linear case all these results emerge as long as $\Psi^{''} < 0$ so that social welfare is concave and there is a concern for redistribution. When $\Psi^{''} = 0$ we return to the first-best solution.

To sum up, while the nonlinear policy brings us back to the first-best Pigouvian rule for the marginal subsidy, it continues to imply a downward distortion on formal care and there will be more traditional couples than efficient. Consequently, the potential conflict between child care provision and redistribution does not solely arise with linear instruments.

Finally, let us revisit the underlying information structure. We have assumed for simplicity that a couple's formal care and career path are observable. We have made this assumption for the ease of exposition, but the arguments and results we presented make clear that the observability of the career path is effectively not necessary. The policy we characterize here can be implemented as long as a couple's level of formal care is observable. This is because high-career couples need full-time care so that their choice of child care would reveal any attempt to mimic a traditional couple. Similarly, a traditional couple mimicking a high-career one would have to choose full-time day care so that mimicking involves the same consumption bundle with or without observable career paths.

The main results of this section are summarized in the following proposition.

Proposition 5 *Assume that couples' formal child care is observable and can be provided publicly at level $c_p^g(q)$ or subject to a nonlinear tax or subsidy. The optimal incentive compatible policy when redistribution is relevant ($\Psi'' < 0$) implies:*

(i) that there is pooling within the traditional and the high career couples groups: all traditional couples receive the same level of formal care and pay the same tax and similarly for all high career couples.

(ii) that high-career couples receive full-time formal care, while the level of c_p^g implies an implicit marginal subsidy which is determined by the Pigouvian rule: it equals $s^g = [1 - F(\hat{q}^g)]\gamma$ which reflects the marginal social damage represented by the norm cost imposed on the high-career couples.

(iii) $\hat{q}^g > \hat{q}^{fb}$ so that there are more traditional couples in the second best than in the first-best. The marginal couple is distorted upwards to reduce the high-career couples' informational rents which improves redistribution.

(iv) $s^g < s^{fb}$; while both levels are set according to the Pigouvian rule, the inequality follows because there are less high-career couples in the second best so that the marginal social damage of the norm cost is smaller.

(v) that, in the GWG, both child penalties and adverse sorting are inefficiently high.

The intuition for (v) is the same as for the corresponding part in Propositions 3 and 4. Again, the policy mitigates the inefficiency of the *laissez-faire* informal care provision but while welfare is obviously higher with the nonlinear policy than with the linear one first-best efficiency is not restored.

7 Social comparisons among peers

In this section we consider a more sophisticated norm based on (local) social comparisons among peers. In other words, the norm that is relevant to any given mother is not determined by the informal average child care in the whole population but by that in her social group. For instance, mothers could refer to the average behavior of female family members or to the behavior of the women they interact with in their everyday life. Groups could also be defined on a geographical basis such as North vs. Southern Italy, where women are exposed to different social environments.[20]

For simplicity assume that there are two social groups, indexed A and B. The groups may differ in their distribution of q, $F^A(q)$ and $F^B(q)$ and in the norm costs γ^A and γ^B. For simplicity and without loss of generality we assume that the two groups have equal size. Preferences are the same as in a one-group model and (apart from the norm cost) are the same for all individuals. Consequently, for $k = A, B$, we have

$$W_{h\ell}^k(q) = y + q + c_p^k y - T_{h\ell}^k + v\left(1 - c_p^k\right) + \beta v\left(c_p^k\right),$$
$$W_{hh}^k(q) = 2y + (1+\alpha)\, q - T_{hh}^k + \beta v\,(1) - \gamma^k F\left(\hat{q}^k\right)\left(1 - c_p^k\right).$$

[20] Social norms on gender roles are much stronger and the gender wage gap is higher in the South of Italy than in the North (Piazzalunga 2018), while policies are the same all over the country.

The optimal allocation is obtained exactly like in Section 3, except that the objective is redefined as $SW = SW^A + SW^B$, where the term pertaining to each group is obtained from (5) by adding the superscript to the relevant variables. Similarly the budget constraint is the sum of expressions in (6). Note that the budget constraint is "global" so that redistribution may occur both within and between groups. We skip all the straightforward algebra but explain the derivation of some expressions which involve some more rearrangements in Appendix (A.7) and (A.8). The FB levels of c_p^k $(k = A, B)$ continue to be given by expression (7) and we have

$$\beta v'(c_p^{fb,k}) + [1 - F^k(\hat{q}^{fb,k})]\gamma^k = v'(1 - c_p^{fb,k}),$$

which are group specific and, if individualized lump-sum transfers $T_{hh}^k(q)$ and $T_{h\ell}^k(q)$ were available, could be implemented by the Pigouvian subsidy s^k given by

$$s^{fb,k} = [1 - F^k(\hat{q}^{fb,k})]\gamma^k.$$

To study second best policies and the robustness of our results we have to distinguish between policies with and without "tagging". Tagging, as defined in the optimal taxation literature, means that the policy (linear or non linear) can be conditioned on the couple's group, which is an exogenous variable.[21] In that case it directly follows that all of our results continue to apply. Specifically, in the linear case we have

$$s^{ok} = \gamma^k \frac{(1 - F^k(\hat{q}^{sk}))E_{hh}^k[\Psi']}{E[\Psi']} - \frac{\mathrm{cov}^k[\Psi', c_p^{sk}]}{E[\Psi']\frac{\partial E^k[c_p^{sk}]}{\partial p^n}}$$

which, except for the superscripts is the same as (20) and we have $s^{ok} < s^{fb}$ as with a single group.[22]

Under nonlinear policies we have,

$$s^{gk} = [1 - F^k(\hat{q}^{gk})]\gamma^k.$$

and the Pigouvian rule applies again. To sum up, the main message of our paper continues to hold when there are several groups and when tagging is possible.

However, tagging is admittedly a strong assumption. In the linear case it is at odds with the assumption that the policy is uniform. More fundamentally, in both the linear and nonlinear cases it requires that groups are observable *and* that it is politically acceptable to use this information to design a policy. In the optimal tax literature tagging is typically considered as problematic (except in special cases) because it may be considered as arbitrary discrimination and violate "horizontal equity" requirements. A more meaningful, and demanding, test for the robustness of the results is thus the case without tagging to which we now turn.

[21] See for instance Cremer *et al.* (2010).

[22] We can say this for sure as long as $E_{hh}^k[\Psi'] < E[\Psi']$; since the latter is not indexed (the budget constraint is global) this might be violated under some assumptions on F^k. But this is just a sufficient condition given that the covariance term is necessarily negative.

7.1 Uniform policy, no tagging

The problem is essentially the same as the linear case with tagging except that we now impose $s^A = s^B = s$ and similarly for τ. Roughly speaking the FOCs are then simply the sum of the FOCs under tagging. We show in Appendix A.7 that the optimal uniform level of s is given by

$$s^o = \gamma^A \frac{(1 - F^A(\hat{q}^{sA}))E^A_{hh}[\Psi']}{E[\Psi']}\delta^A + \gamma^B \frac{(1 - F^B(\hat{q}^{sB}))E^B_{hh}[\Psi']}{E[\Psi']}(1-\delta^A) - \frac{\text{cov}[\Psi', c^s_p]}{E[\Psi']\left(\frac{\partial E^A[c^s_p]}{\partial p^n} + \frac{\partial E^B[c^s_p]}{\partial p^n}\right)}, \tag{33}$$

where

$$\delta^A = \frac{\frac{\partial E^A[c^s_p]}{\partial p^n}}{\frac{\partial E^A[c^s_p]}{\partial p^n} + \frac{\partial E^B[c^s_p]}{\partial p^n}}, \tag{34}$$

so that $0 < \delta^A < 1$. Note that $E[\Psi']$ is defined over both groups; as in Section 5 it is equal to the multiplier associated with the budget constraint which, as explained above, is "global". Note that the amount of formal care purchased by traditional couples is the same in the two groups and satisfies $v'(1 - c^s_p) - s^o = \beta v'(c^s_p)$. However, the *average* amount of formal care differ in the two groups via the distribution $F^k(q)$.

To interpret expression (33) it is better to start with the case where $\Psi'' = 0$ so that redistribution is not a concern. Expression (33) then reduces to

$$s^o = \gamma^A((1 - F^A(\hat{q}^{*A}))\delta^A + \gamma^B((1 - F^B(\hat{q}^{*B}))(1 - \delta^A). \tag{35}$$

When there was a single group (or under tagging), absent redistribution concerns, a linear subsidy set at the (group-specific) Pigouvian level was sufficient to reestablish efficiency. However, with two (or more) groups this is no longer possible because the subsidy must be uniform across groups. The optimal uniform policy then strikes a compromise between the two relevant Pigouvian levels; s^o is set at a weighted average of the group-specific levels.

This effect continues to be present when redistribution matters, as we can see by comparing the first two terms of (33) to the first term of (20). Most significantly from our perspective, the redistributive concern continues to decrease the optimal uniform subsidy. For the reasons explained in Section 5 the covariance term in (33) is still negative and the two first terms in this expressions are smaller than their counterparts in (35) as long as $E^k_{hh}[\Psi'] < E[\Psi']$, for $k = A, B$ so that the average social marginal utility of income is smaller in both high career groups than in the entire population.

This shows that the trade-off between efficiency and redistributive considerations persists with local social comparisons: the optimal subsidy s^o is lower than the Pigouvian level which applies when efficiency is the only social concerns.

7.2 Nonlinear policy, no tagging

When tagging is not possible the optimal policy has to rely on self-selection. The same menu of contracts has to be offered to everyone but the incentive constraints ensure that all couples choose the package $\{T_{h\ell}, c_p^g\}$ designed for them. The menu includes four contracts, $\{T_{h\ell}^A, c_p^{gA}\}$ and $\{T_{h\ell}^B, c_p^{gB}\}$ for $h\ell$-couples and $\{T_{hh}^A, 1\}$ and $\{T_{hh}^B, 1\}$ for hh-couples. Since high career couples of both groups must benefit from full time child care they cannot be separated and incentive compatibility requites $T_{hh}^A = T_{hh}^B = T_{hh}$. The incentive constraint for low career couples requires

$$y + c_p^{gA} y - T_{h\ell}^A + v\left(1 - c_p^{gA}\right) + \beta v\left(c_p^{gA}\right) = y + c_p^{gB} y - T_{h\ell}^B + v\left(1 - c_p^{gB}\right) + \beta v\left(c_p^{gB}\right). \quad (36)$$

In words, low career couples of any given q must received the same utility in both groups. They may receive different levels of child care but the T^k must be adjusted so that they are on the same indifference curve. Note that norm costs are of no relevance for low career couples. In addition the counterpart to equation (23) must be satisfied within each group so that high career couples do not want to mimic low career couples of any of the groups. With (36), this requires, for $k = A, B$,

$$\begin{aligned} &2y + (1+\alpha)\,\hat{q}^{gk} - T_{hh} + \beta v\,(1) - \gamma^k F^k\left(\hat{q}^{gk}\right)\left(1 - c_p^{gk}\right) \\ &- \left(y + \hat{q}^{gk} + c_p^{gk} y - T_{h\ell} + v\left(1 - c_p^{gk}\right) + \beta v\left(c_p^{gk}\right)\right) = 0. \end{aligned} \quad (37)$$

We show in Appendix A.8 that maximizing $SW = SW^A + SW^B$ subject to the global budget constraint and the incentive constraints (36) and (37) yields[23]

$$v'\left(1 - c_p^{gk}\right) - \beta v'\left(c_p^{gk}\right) = (1 - F\left(\hat{q}^{gk}\right))\gamma^k + \gamma^k \frac{E^k[\Psi'] - E[\Psi']}{E[\Psi']}, \; k = A, B. \quad (38)$$

In words, within each group we obtain the Pigouvian rule modified to account for redistribution between groups. Note that with tagging we would have $E^A[\Psi'] = E^B[\Psi'] = E[\Psi']$ so that average marginal social utilities of income would be equalized across groups. In that case the expressions would reduce to the simple Pigouvian rule, as mentioned when tagging was discussed.

Absent of tagging we must have $T_{hh}^A = T_{hh}^B = T_{hh}$ which limits the possibility of redistribution between groups and the solution implies in general different levels of $E^k[\Psi']$. While utilities of low and high career couples for any given q are equalized across groups, the marginal couples differ because of the norm costs and the distribution which are group-specific. Hence, the regressive within group effect of public child care is not relevant. This major message obtained in the one group case is thus robust.

To further interpret expression (38) assume for instance that $E^A[\Psi'] > E[\Psi'] > E^B[\Psi']$ so that group A is "poorer". This yields $c_p^{gA} > c_p^{fb} > c_p^{gB}$ and $s^{gA} > s^{fb} > s^{gB}$. The level of

[23] As part of the proof we show that the Lagrange multiplier associated with (36) is equal to zero. In other words, utility level of low career couples are equalized anyway, even when this is not imposed as extra constraint.

formal child care (and thus the marginal subsidy) is larger than the Pigouvian level in the poor group and lower than this level in the more wealthy group. Thus, child care provision becomes progressive when it comes to redistribution between groups because traditional couples in the poorer group receive a larger level of child care than those in the richer group.

To sum up, with a nonlinear policy there is no conflict between efficient child care provision and redistribution among income levels.

8 Policy discussion

Elizabeth Warren (a democratic candidate for the US presidential elections) has included "universal child care" as a main pillar in her electoral platform. Similarly, in recent Bavarian elections a new "Free voters of Bavaria" movement managed to unsettle the traditional Christian Democratic majority in the regional parliament with a program aiming at offering free child care to all families. Whether or not these are realistic policy options or utopian visions that are impossible to finance (and mainly a boon for well off couples) remains to be seen. But these two examples (which could be completed by many others) show how significant these issues are in practice. Policy choices that are made in the coming years may affect gender roles (and even fertility decisions) for many decades to come.

In most countries the current situation is not the *laissez-faire* allocation used as reference in our analysis. Various policies already provide child care and early childhood education. In the majority of countries, education now begins for most well before 5 years old: 71.5% of young children aged 3 and 4 years are enrolled in education across OECD countries as a whole, and this rises to 79.8% in the OECD countries that are part of the European Union.[24] Publicly-funded pre-primary provision tends to be more strongly developed in the European than in the non-European countries of the OECD. In Europe, the concept of universal access of 3- to 6-year-olds is generally accepted. Most European countries provide all children with at least two years of free, publicly-funded provision before they begin primary school provision. Public expenditure on early childhood and educational care, in cash or in kind, represents today on average 0.8 percent of GDP in OECD countries.[25]

Typically, neither the nursery school nor the primary school provide a form of child care which fully covers the daily needs of full-time working parents.[26] In addition, apart from Scandinavian

[24] Enrolment rates for early childhood education at this age range from over 90% in Belgium, Denmark, France, Germany, Iceland, Italy, New Zealand, Norway, Spain, Sweden and the United Kingdom, at one end of the spectrum, to less than a third in Australia, Greece, Korea, Switzerland and Turkey.

[25] It attains 2 percent in Denmark, and is above 1 percent in the rest of Scandinavia, the United Kingdom, and France. North American and Southern EU countries have the lowest rates of early childhood public spending. In the United States, early childhood public spending is 0.4 percent of GDP; see OECD (2014).

[26] Average hours in early childhood education and care differ substantially across countries. In most OECD

countries, the demand for day-care centers is significantly larger than the available capacity, even in countries with long parental leave. In countries where public funding for such provision is limited, most working parents must either seek (complementary or alternative) solutions in the private market, where ability to pay significantly influences accessibility to quality services, or else rely on informal arrangements with family, friends and neighbors; see OECD (2010).

Fees charged to parents for publicly provided early child-care are often high. Parents in Ireland, the Netherlands, Switzerland and the United Kingdom face some of the highest out-of-pocket costs for centre-based care in Europe. Even though all countries except Ireland provide additional financial support for families on very low incomes, net fees often remain high in absolute terms.[27] This explains, at least to some extent, the fact that children are more likely to use early childhood education and care services when they come from relatively advantaged socio-economic backgrounds; see OECD (2017).

To sum up, currently most child care systems are not designed in such a way to accommodate working parents' needs. The supply of day-care facilities is rationed in terms of spots available, opening hours are generally too short and fees tend to be quite high for children of 0–3 years of age. As long as this remains the case, child care policies notwithstanding, the current situation suffers from the same deficiencies as the laissez-faire in our model. The policies we present, though only second best, would represent a step in the right direction.

Our model shows that a "free for all" approach would be neither efficient nor fair. This would be overshooting in the opposite direction. Our analysis also suggests that attendance can be used as a devise to screen high- and low-income families because the number of hours children spend in day-care represents a proxy for the family's income. Hence, fees contingent on the time children spend in the facility are efficiency enhancing. However, in OECD countries, day-care fees are generally based on enrolment and possibly on family's size and income but to a much lower extent on hours of attendance; see OECD (2017). In addition, information about hours of attendance is typically easier to collect than information about family income or wealth and is not falsifiable. Which again points at children's hours of attendance as a practical screening instrument.

countries, children in early childhood education and care (0-2-year-olds) use it for an average between 25 and 35 hours during a usual week, with the OECD average just under 30 hours per week. However, in some countries (e.g. Iceland, Latvia and Portugal) average hours approach 40 hours during a usual week. In others, such as the Netherlands and the United Kingdom, 0-2 years old in early education centers are there for an average of less than 20 hours during a usual week.

[27] In the Netherlands and the United Kingdom the out-of-pocket cost of full-time centre-based care for two children (aged 2 and 3) in a low-earning dual-earner family works out at around 20% of family disposable income, and at 35% in Ireland.

9 Conclusion

We have studied the design of child-care policies when women's career choices are endogenous. High career mothers suffer from a norm cost caused by "mothers' guilt". Through their child care choices low career mothers create a negative externality via the norm cost. Consequently, the *laissez faire* solution is inefficient; it implies too much informal child care and a share of high-career mothers which is too low.

Child-care policies are effective in enhancing efficiency and reducing gender inequalities. However, since they provide larger benefits to high income couples, they tend to be regressive. Under full information, this effect can be offset by lump-sum transfers associated with a Pigouvian subsidy on formal child care. A uniform subsidy, on the other hand, involves a trade-off between efficiency and redistribution across couples and should be set below the Pigouvian level. Under a nonlinear policy the first-best "Pigouvian" rule for the (marginal) subsidy on formal care is reestablished. While the share of high career mothers continues to be distorted downward for incentive reasons, this policy is effective in reconciling the objectives of reducing the inefficiency in informal care provision and achieving a more equal income distribution across couples.

Our message is robust to a more sophisticated specification of the social norm such that "mothers' guilt" emerges only when women compare their informal care provision to the one of mothers belonging to their reference group. In this case, policies have to address income redistribution both within and across social groups and a linear policy may become even more regressive than in the case of a less sophisticated social norm. Specifically, the group with the larger share of career mothers is likely to be richer and to benefit relatively more from a linear subsidy. With a nonlinear policy instead we obtain the Pigouvian rule within each group modified to account for redistribution between groups. Hence, we show that with a nonlinear policy there is no conflict between efficient formal care provision in each group and redistribution among income levels within and across groups.

From a practical perspective a non-linear policy can be implemented through in-kind provision of child care, at different levels, depending on the mothers career path, and financed with non-linear taxes. Alternatively non-linear subsidies on market care can be used.[28] Either way, day-care fees should be contingent on the amount of time children spend in the facility. More generally, our model indicates that providing "free child care to all" is problematic. While universal provision of preschool child care is desirable, free access is never optimal because it

[28] See also Cremer and Gahvari (1997) who show that when individual consumption levels are observable, in-kind transfers and nonlinear subsidies are equivalent. This information structure also differentiates our model from the extensive literature on in-kind transfers of which child care is of course a prime example (see Blomquist and Chirstiansen, 1995, or Blomquist *et al.,* 2010, for two examples amongst many).

represents a too regressive policy.

Our model has ignored a certain number of important aspects. For instance, we do not consider the welfare of children and the impact of early education on their human capital. There is now ample evidence that high quality formal child care yields better outcome for the children than informal care by less advantaged mothers (see Duncan and Sojourner, 2013, Cornelissen *et al.,* forthcoming). This is likely to call for an even more generous child care policy and tend to increase subsidies. We have also ignored the issue of explicit means testing. This would require a more complex information structure to keep the problem interesting. In essence we would have to combine our approach with a more traditional optimal tax model inspired, for instance, by Casarico *et al.* (2015). These and further extensions are on our research agenda.

References

[1] **Akerlof, G.A., and R.E. Kranton**, "Economics and identity," *The Quarterly Journal of Economics*, 2000, 715–753.

[2] **Akerlof, G.A., and R.E. Kranton**, "Identity economics," *The Economists' Voice*, 2010, *7* (2).

[3] **Barigozzi, F., Cremer H., and K. Roeder**, "Women's Career Choices, Social Norms and Child Care Policies", *Journal of Public Economics*, 168, 162-173, 2018.

[4] **Bernal, R.**, "The effect of maternal employment and child care on children's cognitive development,"*International Economic Review*, 2008, *49* (4).

[5] **Bertrand, M., Kamenica, E. & Pan**, **J**. "Gender identity and relative income within households". *Quarterly Journal of Economics*, 2015, 130, 571–614.

[6] **Blau, F.D., and L.M. Kahn**, "The gender wage gap: extent, trends, and explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

[7] **Blinder, A.**, "Wage discrimination: reduced form and structural estimates," *Journal of Human Resources*, 1973, *8*, 436–55.

[8] **Blomquist, S. and V. Christiansen,** "Public Provision of Private Goods as a Redistributive Device in an Optimum Income Tax Model," *Scandinavian Journal of Economics*, 1995, 97, 547–567.

[9] **Blomquist, S., V. Christiansen and L. Micheletto**, "Public Provision of Private Goods and Nondistortionary Marginal Tax Rates," *American Economic Journal: Economic Policy,* 2010, 2, 1–27.

[10] **Boadway, R., and Keen, M.,** "Public Goods, Self-Selection and Optimal Income Taxation". *International Economic Review*, 1993, 34(3), 463-478.

[11] **Bursztyn L., Fujiwara T. and Pallais A.**, "'Acting wife': marriage market incentives and labor market investments", *American Economic Review*, 2017, *107* (11), 3288–3319.

[12] **Casarico A., L. Micheletto and A. Sommacal**, "Intergenerational transmission of skills during childhood and optimal public policy," 2015, *Journal of Population Economics*, 28 (2), 353–372.

[13] **Card, D., A.R. Cardoso, and P. Kline**, "Bargaining, sorting, and the gender wage gap: quantifying the impact of firms on the relative pay of women." *Quarterly Journal of Economics*, 2016, *131* (2), 633–86.

[14] **Chiappori, P.-A., Salanié, B. & Weiss, Y.** "Partner choice, investment in children, and the marital college premium". *American Economic Review*, 2017, 107, 2109–2167.

[15] **Ciccia, R., and M. Verloo**, "Parental leave regulations and the persistence of the male breadwinner model: using fuzzy-set ideal type analysis to assess gender equality in an enlarged Europe," *Journal of European Social Policy*, 2012, *22* (5), 507–528.

[16] **Cigno, A.** "Fertility and the Tax-Benefit System: A Reconsideration of the Theory of Family Taxation," 1986, *The Economic Journal*, 96(384). 1035–1051.

[17] **Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg.** "Who Benefits from Universal Child Care? Estimating Marginal Returns to early Child Care Attendance." *Journal of Political Economy,* forthcoming.

[18] **Cremer, H., and F. Gahvari, "**In-kind transfers, self-selection and optimal tax policy," *European Economic Review,* 1997, 41, 97–114.

[19] **Cremer H., F. Gahvari, and J.M., Lozachmeur,** "Tagging and Income Taxation: Theory and an Application." *American Economic Journal: Economic Policy*, 2010, 2(1), 31–50.

[20] **Duncan, G.J. and A.J. Sojourner,** "Can intensive early childhood intervention programs eliminate income-based cognitive and achievement gaps?," *Journal of Human Resources,* 2013, 48 (4):945–968.

[21] **Farré, L., and F. Vella**, "The intergenerational transmission of gender role attitudes and its implications for female labour force participation," *Economica*, 2013, *80* (318), 219–247.

[22] **Fortin, N.M.**, "Gender role attitudes and the labour-market outcomes of women across OECD countries," *Oxford Review of Economic Policy*, 2005, *21* (3), 416–438.

[23] **Goldin, C.**, "The quiet revolution that transformed women's employment, education, and family," *The American Economic Review*, 2006, *96* (2), 1–21.

[24] **Goldin, C.** "A grand gender convergence: its last chapter." *American Economic Review*, 2014, *104* (4), 1091–119.

[25] **Guendouzi, J.**, "'The guilt thing': balancing domestic and professional roles," *Journal of Marriage and Family*, 2006, *68* (4), 901–909.

[26] **Huerta, M.d.C., W. Adema, J. Baxter, M. Corak, M. Deding, M.C. Gray, W.-J. Han, and J. Waldfogel**, "Early maternal employment and child development in five OECD countries," *OECD Social, Employment and Migration Working Papers*, 2011, *118*.

[27] **Kleven, H., and C. Landais**, "Gender inequality and economic development: fertility, education and norms", *Economica,* 2017, 84, 180–209

[28] **Kleven, H., C. Landais, and J.E. Søgaard**, "Children and gender inequality: evidence from Denmark", American Economic Journal: Applied Economics, forthcoming, 2018.

[29] **Oaxaca, R.,** "Male–female wage differences in urban labor markets", *International Economic Review*, 1973, *14*, 693–709.

[30] **OECD**, "Doing better for families," *OECD Employment Database*, 2010.

[31] **OECD**, "Education at a Glance", 2010: OECD Indicators, OECD Publishing.

[32] **OECD**, "Starting Strong", 2017: Key OECD Indicators on Early Childhood Education and Care, Starting Strong, OECD Publishing, Paris.

[33] **Olivetti C. and B. Petrongolo**, "The economic consequences of family policies: lessons from a century of legislation in high-income countries" *Journal of Economic Perspective,* 2017, *31* (1), 205–230.

[34] **Paull, G.**, "Children and women's hours of work," *The Economic Journal*, 2008, *118*, F8–F27.

[35] **Piazzalunga, D.** "The Gender Wage Gap among College Graduates in Italy," *Italian Economic Journal*, 2018, 4(1): 33-90.

[36] **Rose, J.**, "Never enough hours in the day: employed mothers' perceptions of time pressure," *Australian Journal of Social Issues*, 2017, *52*, 116–130.

[37] **Rotkirch, A., and K. Janhunen**, "Maternal guilt," *Evolutionary Psychology*, 2010, *8* (1), 90–106.

[38] **Sheshinski, E.**, "The Optimal Linear Income Tax", *Review of Economic Studies*, 1972, 39, 297-302.

Appendix

A.1 Couples' optimization

Below we study couples' decisions focusing on parents' informal care provision. In the main text we focused instead on the complementary choice of formal care purchase.

A.1.1 Only the mother enters the high-career path

Since the mother is in the high-career path, she is not able to take care of the children and suffers the cost of not conforming to the norm. Welfare of the couple is denoted by $W_{\ell h}$ and informal care provided by the father by $c_f = c_{\ell h}$. Noting that $c_{\ell h} + c_p = 1$ and $p = y$, the couple chooses $c_{\ell h}$ to maximize:

$$\max_{c_{\ell h}} \quad W_{\ell h} = y + \alpha q + v(c_{\ell h}) + \beta v(1 - c_{\ell h}) - \gamma \bar{c}_m.$$

Optimal child care provision, $c^*_{\ell h}$, is implicitly determined by:

$$v'(c^*_{\ell h}) = \beta v'(1 - c^*_{h\ell}).$$

Indirect welfare $W^*_{\ell h}$ writes:

$$W^*_{\ell h} = y + \alpha q + v(c^*_{\ell h}) + \beta v(1 - c^*_{\ell h}) - \gamma \bar{c}_m.$$

A.1.2 Both couples enter the low-career path

Here both parents may provide informal care. The social norm for the mother is potentially binding. Welfare of this couple is denoted by $W_{\ell\ell}$ and total informal care provided by the parents is $c_{\ell\ell} = c_m + c_f$. Noting that $c_m + c_f + c_p = c_{\ell\ell} + c_p = 1$ and $p = y$, the couple chooses $c_{\ell\ell}$ to maximize:

$$\begin{aligned}\max_{c_m; c_f} \quad W_{\ell\ell} =& (1 - c_f)\, y + (1 - c_m)\, y - p(1 - c_{\ell\ell}) \\ &+ v(c_{\ell\ell}) + \beta v(1 - c_{\ell\ell}) - \gamma(\max\{0; \bar{c}_m - c_m\}) \\ =& y + v(c_{\ell\ell}) + \beta v(1 - c_{\ell\ell}) - \gamma(\max\{0; \bar{c}_m - c_m\})\end{aligned}$$

Optimal child care provision, $c^*_{\ell\ell} = c^*_m + c^*_f$, is implicitly determined by the two conditions:

$$v'(c^*_f) \leq \beta v'(1 - c^*_{\ell\ell})$$
$$v'(c^*_m) \leq \beta v'(1 - c^*_{\ell\ell}) + I\gamma,$$

where I is an indicator function which takes value 1 when the social norm for mothers is binding, namely when $\bar{c}_m - c_m > 0$, and 0 otherwise. Welfare $W^*_{\ell\ell}$ now is:

$$W^*_{\ell\ell} = y + v(c^*_{\ell\ell}) + \beta v(1 - c^*_{\ell\ell}) - \gamma(\max\{0; \bar{c}_m - c^*_m\}).$$

A.2 The optimal allocation

Denoting λ the Lagrangean multiplier with respect to the budget constraint, the FOCs of (5) with respect to the couples' consumption levels can be rewritten as:

$$\frac{\partial SW}{\partial x_{h\ell}(q)} = \Psi'(W_{h\ell}(q))f(q) - \lambda f(q) = 0 \quad \forall q \leq \hat{q}$$
$$\frac{\partial SW}{\partial x_{hh}(q)} = \Psi'(W_{hh}(q))f(q) - \lambda f(q) = 0 \quad \forall q > \hat{q}.$$

so that:

$$\Psi'(W^{fb}_{hh}(q)) = \Psi'(W^{fb}_{h\ell}(q)) = \lambda \quad \Leftrightarrow \quad W^{fb}_{h\ell}(q) = W^{fb}_{hh}(q) \quad \forall q.$$

Equalizing welfare levels across career paths, we can write:

$$x^{fb}_{h\ell}(q) + v(1 - c^{fb}_p(q)) + \beta v(c^{fb}_p(q)) = x^{fb}_{hh}(q) + \beta v(1) - \gamma F(\hat{q}^{fb})(1 - c^{fb}_p(q)) \quad \forall q. \tag{A.1}$$

We now consider the point-by-point derivative of the social welfare with respect to $c_p(q)$. Given that $c_p(q)$ exerts a negative effect on all $hh-$couples we have:

$$\Psi'(W^{fb}_{h\ell}(c^{fb}_p(q)))\frac{-\partial W^{fb}_{h\ell}(c^{fb}_p(q))}{\partial c^{fb}_p(q)}f(q) + \int_{\hat{q}}^{Q} \Psi'(W^{fb}_{hh}(\varepsilon))\frac{\partial W^{fb}_{hh}(\varepsilon)}{\partial \bar{c}}\frac{-\partial \bar{c}}{\partial c^{fb}_p(q)}f(\varepsilon)\,\mathrm{d}\varepsilon = 0$$

which gives:

$$\Psi'(W^{fb}_{h\ell})\left[v'(1 - c^{fb}_p(q)) - \beta v'(c^{fb}_p(q))\right] f(q) + \int_{\hat{q}}^{Q} \Psi'(W^{fb}_{hh})\left(-\gamma f(q)\right) f(\varepsilon)\,\mathrm{d}\varepsilon = 0$$

Considering that $W^{fb}_{h\ell} = W^{fb}_{hh}$, we can simplify the previous equation as follows:

$$v'(1 - c^{fb}_p(q)) - \beta v'(c^{fb}_p(q)) - \gamma \int_{\hat{q}}^{Q} f(\varepsilon)\,\mathrm{d}\varepsilon = 0$$

showing that it must be $c^{fb}_p(q) = c^{fb}_p$ $\forall q$. Rearranging, the above equation we obtain (27) in the main text.

Taking the derivative of the social welfare function with respect to the marginal couple $\hat{q}$ and rearranging, yields:

$$\alpha\hat{q}f(\hat{q}^{fb}) = f(\hat{q}^{fb})\left[x_{hh}^{fb}(\hat{q}^{fb}) - x_{h\ell}^{fb}(\hat{q}^{fb}) - \gamma[1 - F(\hat{q}^{fb})](1 - c_p^{fb})\right]. \tag{A.2}$$

Given that $c_p(q) = c_p \ \forall q$, we observe that $x_{h\ell}^{fb}(q) = x_{h\ell}^{fb}$ and $x_{hh}^{fb}(q) = x_{hh}^{fb} \ \forall q$. Hence, equation (A.1) can be rewritten as:

$$x_{hh}^{fb} - x_{h\ell}^{fb} = v(1 - c_p^{fb}) + \beta[v(c_p^{fb}) - v(1)] + \gamma F(\hat{q}^{fb})(1 - c_p^{fb}) > 0 \tag{A.3}$$

With (A.3) we can rewrite (A.2) as (9) in the main text.

A.3 Comparative statics

Child care, c_p, and the marginal couple, $\hat{q}$, are implicitly determined by the following two equations:

$$f_1(c_p, \hat{q}, p^n) \equiv y - p^n - v'(1 - c_p) + \beta v'(c_p) = 0$$
$$f_2(c_p, \hat{q}, p^n) \equiv y - \alpha\hat{q} + c_p y + p^n(1 - c_p) + v(1 - c_p) + \beta[v(c_p) - v(1)] + \gamma_m F(\hat{q})(1 - c_p)$$

When we want to know the effect in price changes of formal child care, we have to solve:

$$\begin{bmatrix} \frac{\partial f_1}{\partial c_p} & \frac{\partial f_1}{\partial \hat{q}} \\ \frac{\partial f_2}{\partial c_p} & \frac{\partial f_2}{\partial \hat{q}} \end{bmatrix} \begin{bmatrix} \mathrm{d}c_p \\ \mathrm{d}\hat{q} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1}{\partial p^n} \\ \frac{\partial f_2}{\partial p^n} \end{bmatrix} \mathrm{d}p^n.$$

Inserting the derivatives and inverting the first matrix, we have:

$$\begin{bmatrix} \mathrm{d}c_p \\ \mathrm{d}\hat{q} \end{bmatrix} = \frac{1}{D} \begin{bmatrix} -\alpha + \gamma f(\hat{q})(1 - c_p) & 0 \\ \gamma F(\hat{q}) & v''(1 - c_p) + \beta v''(c_p) \end{bmatrix} \begin{bmatrix} 1 \\ -(1 - c_p) \end{bmatrix} \mathrm{d}p^n,$$

where $D = [-\alpha + \gamma f(\hat{q})(1 - c_p)][v''(1 - c_p) + \beta v''(c_p)] > 0$. We thus have:

$$\frac{\mathrm{d}c_p}{\mathrm{d}p^n} = \frac{1}{v''(1 - c_p) + \beta v''(c_p)} < 0 \tag{A.4}$$

$$\frac{\mathrm{d}\hat{q}}{\mathrm{d}p^n} = \frac{-[v''(1 - c_p) + \beta v''(c_p)](1 - c_p) + \gamma F(\hat{q})}{[-\alpha + \gamma f(\hat{q})(1 - c_p)][v''(1 - c_p) + \beta v''(c_p)]} > 0 \tag{A.5}$$

A.4 Uniform subsidies

The FOC wrt s can be written as

$$E[\Psi' c_p^s] + (1 - F(\hat{q}^s))E_{hh}[\Psi']\gamma\left[F(\hat{q}^s)\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} - (1 - c_p^s)f(\hat{q}^s)\frac{\mathrm{d}\hat{q}^s}{\mathrm{d}p^n}\right] - E[\Psi']E[c_p^s]$$
$$- E[\Psi']s\left[-F(\hat{q}^s)\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} + (1 - c_p^s)f(\hat{q}^s)\frac{\mathrm{d}\hat{q}^s}{\mathrm{d}p^n}\right] = 0,$$

where $E[c_p^s] = F(\hat{q}(p^n))c_p(p^n) + 1 - F(\hat{q}(p^n))$. Noting that

$$\frac{\partial E[c_p^s]}{\partial p^n} = F(\hat{q}^s)\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} - (1 - c_p^s)f(\hat{q}^s)\frac{\mathrm{d}\hat{q}^s}{\mathrm{d}p^n} < 0 \tag{A.6}$$

and $\mathrm{cov}[\Psi', c_p^s] = E[\Psi' c_p^s] - E[\Psi']E[c_p^s]$, we can write

$$\frac{\partial SW}{\partial s} = \mathrm{cov}[\Psi', c_p^s] - (1 - F(\hat{q}^s))E_{hh}[\Psi']\gamma\frac{\partial E[c_p^s]}{\partial p^n} + E[\Psi']s\frac{\partial E[c_p^s]}{\partial p^n}. \tag{A.7}$$

Setting this expression equal to zero and solving for s yields equation (20). Further evaluating (A.7) at the Pigouvian level $s^{fb} = [1 - F(\hat{q}^{fb})]\gamma_m$, which from (13) implies $\hat{q}^* = \hat{q}^{fb}$ yields

$$\left.\frac{\partial SW}{\partial s}\right|_{s=s^{fb}} = \mathrm{cov}[\Psi', c_p^s] - (1-F(\hat{q}^{fb}))E_{hh}[\Psi']\gamma\frac{\partial E[c_p^s]}{\partial p^n} + E[\Psi'][1-F(\hat{q}^{fb})]\gamma_m\frac{\partial E[c_p^s]}{\partial p^n} = \mathrm{cov}[\Psi', c_p^s] < 0 \tag{A.8}$$

so that assuming concavity we must have $s^o < s^{fb}$.

A.5 Proof of equation (28)

The FOC wrt c_p^g is given by:

$$\int_0^{\hat{q}^g} \Psi'(\cdot)\left[y - v'\left(1 - c_p^g\right) + \beta v'\left(c_p^g\right)\right] f(q)\,\mathrm{d}q + \int_{\hat{q}^g}^{Q} \Psi'(\cdot)\,\gamma F(\hat{q}^g)\, f(q)\mathrm{d}q$$
$$- \widehat{\lambda} p F(\hat{q}^g) + \mu\left[\gamma F(\hat{q}^g) - y + v'\left(1 - c_p^g\right) - \beta v'\left(c_p^g\right)\right] = 0.$$

With equations (25) and (26) and the following definitions:

$$E_{h\ell}[\Psi'] = \frac{\int_0^{\hat{q}^g} \Psi'(\cdot)f(q)\mathrm{d}q}{F(\hat{q}^g)} \quad \text{and} \quad E_{hh}[\Psi'] = \frac{\int_{\hat{q}^g}^{Q} \Psi'(\cdot)f(q)\mathrm{d}q}{1 - F(\hat{q}^g)}$$

we can rewrite the above FOC as:

$$E_{h\ell}[\Psi']F(\hat{q}^g)[y - v'(1 - c_p^g) + \beta v'(c_p^g)] + \gamma F(\hat{q}^g)E_{hh}[\Psi'](1 - F(\hat{q}^g)) - E[\Psi']yF(\hat{q}^g)$$
$$+ [-E_{hh}[\Psi'](1 - F(\hat{q}^g)) + E[\Psi'](1 - F(\hat{q}^g))]\left[\gamma F(\hat{q}^g) - y + v'\left(1 - c_p^g\right) - \beta v'\left(c_p^g\right)\right] = 0. \tag{A.9}$$

Noting that $E_{h\ell}[\Psi']F(\hat{q}^g) + E_{hh}[\Psi'](1 - F(\hat{q}^g)) = E[\Psi']$, we can write:

$$E[\Psi'][y - v'(1 - c_p^g) + \beta v'(c_p^g)] - E[\Psi']yF(\hat{q}^g)$$
$$+ E[\Psi'](1 - F(\hat{q}^g))[\gamma F(\hat{q}^g) - y + v'\left(1 - c_p^g\right) - \beta v'\left(c_p^g\right)] = 0$$

which reduces to:

$$[1 - F(\hat{q}^g)]\gamma - v'(1 - c_p^g) + \beta v'(c_p^g) = 0.$$

A.6 Proof of equation (31)

Solving the IC constraint for $T_{h\ell} - T_{hh}$ yields

$$T_{h\ell} - T_{hh} = -y - \alpha\hat{q}^g - \beta v\left(1\right) + \gamma F\left(\hat{q}^g\right)\left(1 - c_p^g\right) + c_p^g y + v\left(1 - c_p^g\right) + \beta v\left(c_p^g\right)$$

From (13) we have the first-best marginal couple:

$$\hat{q}^{fb} \equiv \frac{1}{\alpha}\left[v(1 - c_p^{fb}) + \beta v(c_p^{fb}) - \beta v(1) + \gamma F(\hat{q}^{fb})(1 - c_p^{fb}) - \gamma[1 - F(\hat{q}^{fb})](1 - c_p^{fb})\right]$$

We now substitute $c_p^g = c_p^{fb}$ and $\hat{q}^g = \hat{q}^{fb}$:

$$\begin{aligned} T_{h\ell} - T_{hh} &= -y - v(1 - c_p^g) - \beta v(c_p^g) + \beta v(1) - \gamma F(\hat{q}^{fb})(1 - c_p^g) + \gamma[1 - F(\hat{q}^{fb})](1 - c_p^g) \\ &- \beta v\left(1\right) + \gamma F(\hat{q}^{fb})\left(1 - c_p^g\right) + c_p^g y + v\left(1 - c_p^g\right) + \beta v\left(c_p^g\right). \end{aligned}$$

The above equation simplifies to:

$$T_{h\ell} - T_{hh} = \gamma[1 - F(\hat{q}^{fb})](1 - c_p^g) - y(1 - c_p^g).$$

A.7 Proof of equation (33)

The objective function is obtained from (16) by indexing all the group-specific variables and summing over $k = A, B$. Similarly, the budget constraint is[29]

$$2\tau = \sum_{k=A}^{B}\left\{sF(\hat{q}^{sk}(p^n))c_p(p^n) + s[1 - F(\hat{q}^{sk}(p^n))]\right\}. \tag{A.10}$$

The FOC with respect to s is then given by

$$\begin{aligned} &\sum_{k=A}^{B}\left\{F(\hat{q}^{sk})E_{h\ell}^k[\Psi']c_p^s + (1 - F(\hat{q}^{sk}))E_{hh}^k[\Psi']\left[1 - \gamma^k F(\hat{q}^{sk})\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} + \gamma^k(1 - c_p^s)f(\hat{q}^{sk})\frac{\mathrm{d}\hat{q}^{sk}}{\mathrm{d}p^n}\right]\right\} \\ &- \lambda\sum_{k=A}^{B}\left[F(\hat{q}^{sk})c_p^s - sF(\hat{q}^{sk})\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} + s(1 - c_p^s)f(\hat{q}^{sk})\frac{\mathrm{d}\hat{q}^s}{\mathrm{d}p^n} + 1 - F(\hat{q}^{sk})\right] = 0. \end{aligned} \tag{A.11}$$

The FOC with respect to τ is given by[30]

$$2\lambda = \sum_{k=A}^{B}\int_0^{Q^k}\Psi'(q)f^k(q)\mathrm{d}q \equiv 2E[\Psi'], \tag{A.12}$$

[29] To keep the model (and the expressions) as close as possible to the one group case, we assume that population "size" in each group is one. Consequently, all the within group expectations defined above remain valid. Total population size is then 2 and the global expectation and coviance is obtained by dividing the usual integral by 2.

[30] Total population size is now equal to 2 so that

$$E[\Psi'] = \frac{1}{2}\sum_{k=A}^{B}\int_0^{Q^k}\Psi'(q)f^k(q)\mathrm{d}q.$$

The expections $E[\Psi', c_p^s]$ and $E[c_p^s]$ are redefined in the same way.

Proceeding like in Appendix A.4, and making use of (A.12) we can rewrite (A.11)

$$E[\Psi', c_p^s] + \sum_{k=A}^{B} (1 - F^k(\hat{q}^{sk}))E_{hh}^k[\Psi']\gamma^k \left[F^k(\hat{q}^{sk})\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} - (1 - c_p^s)f^k(\hat{q}^{sk})\frac{\mathrm{d}\hat{q}^{sk}}{\mathrm{d}p^n}\right] - E[\Psi']E[c_p^s]$$
$$- E[\Psi']s\sum_{k=A}^{B}\left[-F^k(\hat{q}^{sk})\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} + (1 - c_p^s)f^k(\hat{q}^{sk})\frac{\mathrm{d}\hat{q}^{sk}}{\mathrm{d}p^n}\right] = 0, \qquad \text{(A.13)}$$

where $E[c_p^s] = \sum_{k=A}^{B} E^k[c_p^s]/2 = \sum_{k=A}^{B}[F^k(\hat{q}^{sk}(p^n))c_p(p^n) + 1 - F^k(\hat{q}^{sk}(p^n))]/2$. So that equation (A.6) becomes

$$\frac{\partial E^k[c_p^s]}{\partial p^n} = F^k(\hat{q}^{sk})\frac{\mathrm{d}c_p^s}{\mathrm{d}p^n} - (1 - c_p^s)f^k(\hat{q}^{sk})\frac{\mathrm{d}\hat{q}^{sk}}{\mathrm{d}p^n} < 0 \qquad \text{(A.14)}$$

and (A.13) can be written as

$$\mathrm{cov}[\Psi', c_p^s] - \sum_{k=A}^{B}(1 - F^k(\hat{q}^{sk}))E_{hh}^k[\Psi']\gamma^k\frac{\partial E^k[c_p^s]}{\partial p^n} + E[\Psi']s\sum_{k=A}^{B}\frac{\partial E^k[c_p^s]}{\partial p^n} = 0. \qquad \text{(A.15)}$$

Solving for s and rearranging then yields expression (33).

A.8 Proof of expression (38)

First, observe that equation (36) does not impose any extra restriction. This is because when the constraint (36) is *not* imposed the solution implies that utilities of the low career couples are equal across groups for any given q. In other words the constraint is redundant. To see this assume, by constrast, that the solution implies for instance $W_{h\ell}^A(q) < W_{h\ell}^B(q)$. Then, $W_{h\ell}^A(q)$ is not relevant for the IC constraint of any of the high career couples; when an hh couple mimics a $h\ell$ couples it will always claim to be of type B. But then we can slightly decrease the tax by $h\ell$ couples in group A and increase the taxes equally in all other groups to maintain budget balance. This will not affect the incentive constrains and since $h\ell$ couples in group A are the ones with the lowest utility, it is plain that social welfare increases.

Proceeding like in Appendix A.5 we can show that the FOC with respect to c_p^{gk} is given by

$$\int_0^{\hat{q}^{gk}} \Psi'(\cdot)\left[y - v'\left(1 - c_p^{gk}\right) + \beta v'\left(c_p^{gk}\right)\right] f^k(q)\,\mathrm{d}q + \int_{\hat{q}^{gk}}^{Q^k} \Psi'(\cdot)\gamma F^k\left(\hat{q}^{gk}\right) f^k(q)\mathrm{d}q$$
$$- \widehat{\lambda}pF^k\left(\hat{q}^{gk}\right) + \mu^k\left[\gamma^k F^k\left(\hat{q}^{gk}\right) - y + v'\left(1 - c_p^{gk}\right) - \beta v'\left(c_p^{gk}\right)\right] = 0. \qquad \text{(A.16)}$$

while the FOCs with respect to the transfers $T_{h\ell}^k$, $k = A, B$, and $T_{hh}^A = T_{hh}^B = T_{hh}$ are:

$$-E_{h\ell}^k[\Psi']F^k\left(\hat{q}^{gk}\right) + \mu^k + \widehat{\lambda}F^k(\hat{q}^{gk}) = 0 \qquad \text{(A.17)}$$

$$-E_{hh}^A[\Psi']\left[1 - F^A\left(\hat{q}^{gA}\right)\right] - \mu^A + \widehat{\lambda}\left[1 - F^A\left(\hat{q}^{gA}\right)\right]$$
$$-E_{hh}^B[\Psi']\left[1 - F^B\left(\hat{q}^{gB}\right)\right] - \mu^B + \widehat{\lambda}\left[1 - F^B\left(\hat{q}^{gB}\right)\right] = 0 \qquad \text{(A.18)}$$

Combining (A.17) for $k = A, B$ with (A.18) yields

$$\widehat{\lambda} = \sum_{k=A}^{B} \left\{ E^k_{h\ell}[\Psi']F^k\left(\hat{q}^g\right) + E^k_{hh}[\Psi']\left[1 - F^k\left(\hat{q}^{gk}\right)\right]\right\} = E[\Psi'] \tag{A.19}$$

Substituting for $\widehat{\lambda}$ and μ^k from (A.19) and (A.17) into (A.16) then yields

$$E^k_{h\ell}[\Psi']F^k(\hat{q}^{gk})[y - v'(1 - c_p^{gk}) + \beta v'(c_p^{gk})] + \gamma^k F^k(\hat{q}^{gk})E^A_{hh}[\Psi'](1 - F^k(\hat{q}^{gk})) - E[\Psi']yF^k(\hat{q}^{gk})$$
$$+ \left[E^k_{h\ell}[\Psi']F^k\left(\hat{q}^{gk}\right) - E[\Psi']F(\hat{q}^{gk})\right]\left[\gamma^k F\left(\hat{q}^{gk}\right) - y + v'\left(1 - c_p^{gk}\right) - \beta v'\left(c_p^{gk}\right)\right] = 0.$$

Simplifying and rearranging then successively yields

$$\gamma^k F(\hat{q}^{gk})E^k_{hh}[\Psi'](1 - F^k(\hat{q}^{gk})) + \gamma^k F^k(\hat{q}^{gk})E^k_{h\ell}[\Psi']F^k\left(\hat{q}^{gk}\right) - E[\Psi']yF^k(\hat{q}^{gk})$$
$$+ [-E[\Psi']F^k(\hat{q}^{gk})]\left[\gamma^k F\left(\hat{q}^{gk}\right) - y + v'\left(1 - c_p^{gk}\right) - \beta v'\left(c_p^{gk}\right)\right] = 0,$$

$$\gamma^k F^k(\hat{q}^{gk})E^k[\Psi'] - E[\Psi']yF^k(\hat{q}^{gk})$$
$$+ [-E[\Psi']F^k(\hat{q}^g)]\left[\gamma^k F^k\left(\hat{q}^{gk}\right) - y + v'\left(1 - c_p^{gk}\right) - \beta v'\left(c_p^{gk}\right)\right] = 0,$$

$$\gamma^k E^k[\Psi'] - E[\Psi']y$$
$$+ [-E[\Psi']]\left[\gamma^k F^k\left(\hat{q}^{gk}\right) - y + v'\left(1 - c_p^{gk}\right) - \beta v'\left(c_p^{gk}\right)\right] = 0,$$

$$\gamma^k E^k[\Psi']$$
$$+ [-E[\Psi']]\left[\gamma^k F^k\left(\hat{q}^{gk}\right) + v'\left(1 - c_p^{gk}\right) - \beta v'\left(c_p^{gk}\right)\right] = 0,$$

$$\gamma^k E^k[\Psi'] - \gamma^k E[\Psi']$$
$$+ [-E[\Psi']]\left[\gamma^k F^k\left(\hat{q}^{gk}\right) - \gamma^k + v'\left(1 - c_p^{gk}\right) - \beta v'\left(c_p^{gk}\right)\right] = 0,$$

and finally

$$v'\left(1 - c_p^{gk}\right) - \beta v'\left(c_p^{gk}\right) = (1 - F^k\left(\hat{q}^{gk}\right))\gamma^k + \gamma^k \frac{E^k[\Psi'] - E[\Psi']}{E[\Psi']},$$

which is (38).