INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Efficient Child Care Subsidies

Christine Ho, Nicola Pavoni

Working Paper n. 572

This Version: March 8 , 2016

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Efficient Child Care Subsidies*

Christine Ho† Nicola Pavoni‡

March 8, 2016

Abstract

We study the design of child care subsidies in an optimal welfare and tax problem. The optimal subsidy schedule is qualitatively similar to the existing US scheme. Efficiency mandates a subsidy on formal child care costs for working mothers, with higher subsidies paid to lower income earners. The optimal subsidy is also kinked as a function of child care expenditure. To counterbalance the sliding scale pattern of the optimal subsidy rates, marginal labor income tax rates are set lower than the labor wedges, with the potential to generate negative marginal tax rates. We calibrate our model to features of the US labor market and focus on single mothers with children aged below 6. The optimal program provides stronger participation incentives compared to the US scheme. The intensive margin incentives provided by the efficient program are milder, with subsidy rates decreasing with income more steeply than those in the US.

JEL: D82, H21, H24, J13

Keywords: optimal taxation, asymmetric information, child care subsidies.

*We would like to thank Pierre André Chiappori, Stéphane Gauthier, Tullio Jappelli, Jeremy Lise, Emmanuel Saez, Kjetil Storesletten, and attendees at the GRIPS, PET, SES and RES conferences, at the MOVE-FINet-CEAR Workshop on Family Economics, the SOE SMU Research Workshop, and the Macroeconomic and Policy Implications of Underground Economy and Tax Evasion conference.

†School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore. Tel: +65 68085173. Email: christineho@smu.edu.sg.

‡Bocconi University, IGIER, IFS and CEPR. Department of Economics, Universita' Bocconi, via Roentgen 1, 20136 Milano, Italy. Email: nicola.pavoni@unibocconi.it .

1 Introduction

The transition of mothers' role from traditional full-time homemakers to potential breadwinners over the past decades indicates the increasing involvement of mothers as active members of the labor force. In parallel, policy makers are increasing their focus on child care subsidy programs. In the US, programs such as the Dependent Care Tax Credit (DCTC) and the Child Care and Development Fund (CCDF) are benefiting from increased funding.[1] The focus of policy debates has so far been on affordability and quality of child care. As such, the literature on child care subsidy programs has outlined the use of child care subsidies as a tool to promote economic self-sufficiency among low income families and decrease their reliance on welfare.[2]

Even though there is a vast literature on the impact of child care subsidies on employment of mothers and considerable policy debates on affordability of child care, none has so far looked at the optimal design of child care subsidies. We study the design of such subsidies within an optimal welfare and tax problem where agents have private information on labor market productivities. Agents have child care needs and allocate effort between the primary labor market and household child care activities.

We show that it is optimal to pay a positive child care subsidy on formal child care costs and that higher child care subsidies should be paid to lower income earners. We therefore offer an efficiency reason to existing debates for providing child care subsidies to low income earners and suggest that a *sliding scale child care subsidy scheme* would be an optimal way of promoting employment while achieving re-distributional goals. Moreover, very much in line with the qualitative features of the existing scheme in the US, *the optimal subsidy must be kinked* as a function of child care expenditure. An agent whose formal child care expenditure are lower than the kink-point faces a positive subsidy while it is optimal to set a non positive subsidy for child care expenditure above the kink-point.

By jointly designing child care subsidies and income taxation (in the form of income depen-

[1] In 2010, \$3.4bn were made available via the DCTC while in 2013, the CCDF made \$5.3bn available. Recent debates include the 2011 Obama Administration's proposal to double the DCTC for families earning below \$85k (Tax Policy Center, 2010b) and the FY2015 budget requesting an increase of \$807m to fund the CCDF (National Center for Infant, Toddlers, and Families, 2015).

[2] There is a wide array of literature providing evidence of positive impacts of child care subsidies on the labor supply of mothers (Bainbridge, Meyers, and Waldfogel, 2003; Blau, 2003; Blau and Tekin, 2007; Ho, 2013; Ho, 2015b; Kimmel, 1995; Tekin, 2005). In addition, early childhood intervention proponents are providing increasing evidence of the positive benefits of high quality child care on children's outcomes (Karoly et al., 1998; Currie, 2001; Heckman, 2006).

dent child allowances), we show that the new policy tool cannot be replicated by a negative marginal tax rate based on earned income of low skilled workers alone (such as, e.g., the Earned Income Tax Credit in the US). Our implementation exercise, however, generates an interesting discrepancy between the standard labor wedge (which is always positive in our model) and the marginal tax on earned income. In particular, the optimal marginal taxes (inclusive of the income dependent child allowances) are set at lower rates than the labor wedges due to the interaction with the sliding scale pattern of child care subsidies. This discrepancy is particularly relevant at low income levels and may potentially lead to negative marginal taxes on income.

This paper also provides quantitative estimates of the optimal child care subsidy rates. We calibrate our model to features of the US labor market and focus on single mothers with children aged below 6. According to US Census data, the number of children living in single parent homes has nearly doubled between 1960 to 2010 with nearly one third (15 million) of children currently living with a single mother. We chose to focus on single mothers with young children because they tend to have high child care needs and are often targeted by generous transfer programs. Our study is therefore designed to focus on low and middle income earners. For the purpose of this study, we can abstract from the practical complexity of modeling intra-household decisions in two parent households within an optimal tax framework.

We use data from the Current Population Survey (CPS) to calibrate the empirical distribution of market productivities as well as our preference parameters. The presence of child care needs means that we also model labor supply at the extensive margin (in addition to the intensive margin). Given the current tax and transfer system, a proportion of low market productivity agents self-select into unemployment while one may want them to work in the optimal program. We therefore impute the potential wage distribution of unemployed mothers in line with the empirical labor literature. Optimal subsidy rates decrease with income more steeply than those in the current US scheme while optimal child allowances are flatter than those in the US. In the benchmark calibration, the optimal coverage varies from 80% of formal child care cost for single mothers earning below \$10k to 20% for mothers earning approximately \$20k a year. No child care subsidy is paid for labor earnings above \$25k-\$30k. Optimal marginal income tax rates are positive at all earnings levels. The optimal program provides stronger participation incentives but milder intensive margin incentives compared to the US scheme.

Literature Barnett (1993) and Domeij and Klein (2013) argue that child care subsidies should be offered to mothers with young children to counteract the disincentive effects of the current tax system on labor supply.[3] We find that the optimal pattern of child care subsidies across income groups do not mimic at all (neither quantitatively nor qualitatively) the shape of the labor income taxes, suggesting a richer role for such instrument in this context.[4]

To implement the constrained efficient allocation, we allow the government to use child care subsidies on formal child care cost to indirectly tax home activities, which would otherwise be detrimental for incentive compatibility. This is in a similar spirit to the exercises performed in the New Dynamic Public Finance literature (Golosov et al., 2013; Kocherlakota, 2010; Saez, 2002b; Werning, 2011) where both labor supply and saving wedges are considered. The child care margin is different from the saving margin studied in these works, both economically and technically.[5]

The introduction of child care relates our paper to the literature on income taxation in the presence of non-market activities (e.g., Beaudry, Blackorby, and Szalay, 2009; Choné and Laroque, 2011; Saez, 2002a). This literature considers heterogeneous cost of labor market participation and has argued that it is optimal to subsidize low income earners in the form of a negative marginal income tax rate. We consider a different framework where mothers differ in labor market productivities but face the same hourly cost of formal child care. As in these works, our model involves a multidimensional choice problem.[6] Although we are unable to adopt the standard 'local approach', the model permits a sharp characterization of the optimal allocation by focusing on only the downward incentive constraints.

[3]A similar principle emerges in the representative agent model of Kleven, Richter, and Sørensen (2000), who study linear commodity taxation in presence of home production.

[4]In fact, even in the existing US scheme, child care subsidies seem to follow a somewhat more complex pattern. For example, since the Earned Income Tax Credit scheme implies a negative income tax rate for low income earners with young children, if child care subsidies were to merely mimic (counteract) the pattern of the marginal income taxes, child care costs should be taxed - not subsidized - for low income earners.

[5]For example, due to the non-separability between labor supply and child care the implementation of the second best allocation in our model requires a kink in the subsidy schedule. Thanks to the additive separability assumption between consumption and leisure in these studies, savings can instead be taxed linearly. For the need of a kink in savings taxation in the presence of nonseparabilities, see Kocherlakota (2004).

[6]There are important differences in the framework considered, that imply different technical difficulties and require a different approach. In Beaudry, Blackorby, and Szalay (2009), the different activities are perfect substitutes, while in Choné and Laroque (2011) and Saez (2002a), agents face heterogeneous fixed costs of participation to the labor market. Our model contemplates two genuinely different intensive margins (work and child care). Our framework is more closely related to Besley and Coate (1995), but the characteristic of our model does not allow for the (more standard) local-approach adopted in that paper. Instead, we follow a line of attack to the problem that is similar to that indicated by Matthews and Moore (1987).

Also related to our paper is the literature in quantitative macroeconomics that aims at numerically computing welfare gains from policy reforms as opposed to characterizing the fully optimal tax and subsidy scheme as we do. Representative references include Domeij and Klein (2013) and Guner, Ventura, and Kaygusuz (2013). Our work complements these studies in that it analyzes a richer (and hence more flexible) policy tool in a simpler set up. Flexibility supported by rigorous economic principles might be valuable when the aim is to assess the optimality of a complex scheme such as the existing one in the US (see below). Moreover, studying the efficient design of child care subsidies jointly with optimal child allowances allows us to understand how they have an independent role from income taxes.

We document the main components of child care subsidy programs in the US in Section 2. In Section 3, we present our model of the household where mothers choose both labor supply in the primary market and household-provided child care. Optimal policy and implementation results are presented in Sections 4 and 5, respectively. The calibration exercise and numerical results are presented in Section 6. Section 7 concludes.

2 US Child Related Subsidy Programs

In this Section, we describe the 2010 US tax and subsidy scheme, with a particular focus on child care subsidies and child dependent allowances. We outline the main features of interest in two major child care (price related) subsidy programs in the US, the Dependent Care Tax Credit (DCTC) and the Child Care and Development Fund (CCDF). We then describe the child dependent tax exemptions and allowances that are available to families with children under the federal income tax scheme, the Earned Income Tax Credit (EITC), and the Temporary Assistance to Needy Families (TANF). Further details regarding the US welfare program are reported in Appendix B.3.

Child Care Subsidies (DCTC and CCDF) The DCTC is a non-refundable federal income tax credit program available to families with children aged under 13 and covers part of child care expenses. The CCDF is a block grant fund managed by states within certain federal guidelines. CCDF subsidies are available as vouchers or as part of direct purchase programs to families with children under 13 and with income below 85% of the state median income.

Figure 1: 2010 US Tax and Subsidy Schedules

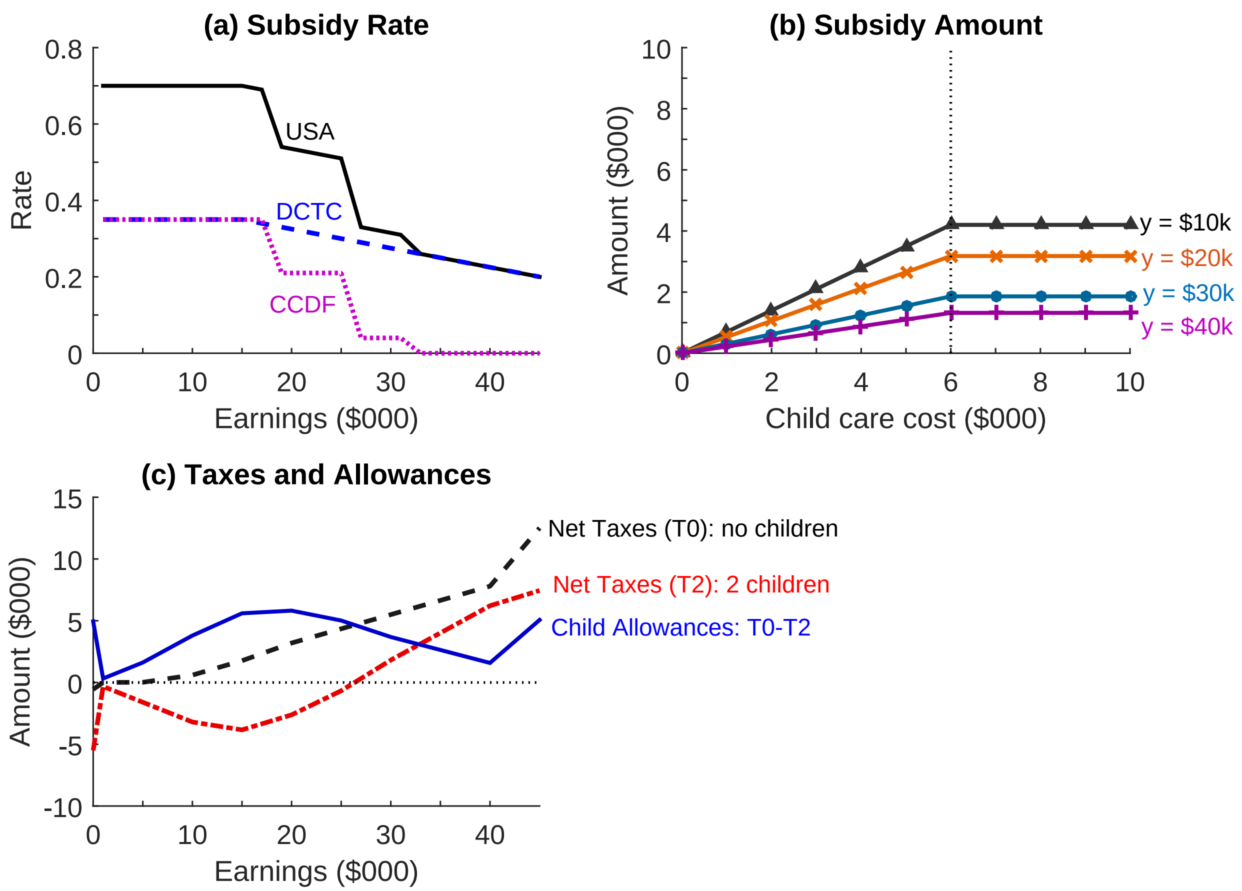

Figure 1: Panel (a) reports child care subsidy rates under DCTC and CCDF, and the consolidated rates (solid line) as a function of gross family income. Panel (b) reports the amounts of child care subsidies received as a function of total formal child care costs and by family income (y). We report the different schedules faced by individuals earnings between $10k and $40k a year. For all income levels, the subsidy rate drops to zero for total child care costs above $6k. All reported schedules are for a family with two children aged below 13. Panel (c) depicts the amounts of net income taxes payable as a function of gross family income for a single person with no children and for a single person with two children. The net income taxes include TANF benefits, federal and social security taxes, and EITC. The difference between net income taxes for a single person without and with children are represented by the solid line, and are interpreted as the child allowances that a parent is eligible for under the US welfare system.

Employment Requirements. Both child care subsidy programs are conditional on employment of parents. In particular, the DCTC is a tax credit available only to families who earn income and pay taxes while the CCDF is available to low income families who are engaged in work related activities.[7]

Sliding Scale. In both the DCTC and the CCDF, the child care subsidy rate declines as income increases.[8] In particular, the DCTC has a tax credit rate of 35% of child care expenses for families with annual gross income of less than $15,000. The tax credit rate declines by 1% for each $2,000 of additional income until it reaches a constant tax credit rate of 20% for families with annual gross income above $43,000. Whereas the Federal recommended subsidy rate for the CCDF is 90%, only a certain proportion of eligible households receive the subsidy: 39%, 24%, and 5% of households living, respectively, below, between 101% and 150%, and above 150% of the poverty threshold received the CCDF subsidy (US Department of Health and Human Services, 2009).[9] Panel (a) of Figure 1 illustrates the average child care subsidy rates under the DCTC and the CCDF according to family income.[10]

Decreasing Coverage. The coverage rate decreases with total expenditure on child care. The DCTC has a cap on child care expenditure of $3,000 for families with one child and $6,000 for families with two children. As of 2010, the CCDF maximum reimbursement rates ranged from $280 per week (Puerto Rico) to $1,465 per month (New York) for an infant in full time formal child care (Minton et al., 2012). In addition, 17 states had a cap on the number of hours of formal child care use, ranging from 45 hours per week (Michigan) to 20 hours per day (Montana). Panel (b) of Figure 1 illustrates the amount of child care subsidy that a family with two eligible children would receive under the DCTC and CCDF, taking the DCTC cap of $6,000 into account. We illustrate the scheme for families with two children as our sample of interest (single mothers with children aged below 6) have two children on average (see Section 6 for details on our sample from the CPS). Consistent with the rates reported in Panel (a), the slope of the

[7] In 2010, 81% of families receiving CCDF were employed, with the remaining families in training (Administration for Children and Families, 2012).

[8] While there are differences across states in the generosity of the subsidy rates, in all states, the child care subsidy rates strictly follow a sliding scale pattern (Gabe, Lyke, and Spar, 2001).

[9] According to federal guidelines, states using CCDF funding are also required to have co-payments from the family that increase with family income. We do not take into account the state wide variations in co-payments in our analysis and focus on the average subsidy rates at the federal level.

[10] Following the allocation rates described above, Figure 1 is drawn by imputing an average CCDF subsidy rate of 35.1%, 21.6%, and 4.5% to households with income below, between 101% and 150%, and above 150% of the poverty threshold, respectively.

subsidy amount schedule before the cap decreases with family income.

Child Allowances (Tax Exemptions, EITC and TANF) In addition to subsidies on the cost of formal child care, parents in the US are also eligible for relatively generous child dependent allowances that are conditional on the presence of children in the household. Under the federal income taxation scheme, taxable income is based on earnings minus standard deductions of $5,700 for a single childless person and $8,400 for a single parent, minus exemptions of $3,650 for each taxpayer and dependent. Both childless individuals and parents are subject to social security (SS) taxes set at 7.65% of earnings. Working families are eligible for the EITC, which is a refundable tax credit and follows a 'trapezoid' pattern.[11]

Parents are also eligible for TANF, which is a cash assistance program for families with children aged below 18. In 2010, nearly 80% of TANF recipients were unemployed while a family with two children received on average $412 of TANF benefits per month (for details, see US Department of Health and Human Services, 2011). We do not explicitly set unemployment insurance benefits as young mothers may not be eligible for them if they have no previous work experience (see Section 6 for details).

Panel (c) of Figure 1 illustrates the net income taxes payable by a single childless person and by a single parent with two children, computed as federal income and SS taxes minus EITC benefits for the employed, and minus TANF and additional benefits for the unemployed. The demographic dependent child allowances are computed as the difference between net taxes of a childless individual and net taxes of a single parent with two children. This figure illustrates at least three qualitative properties of the US tax and transfers system. First, child allowances are by all means equivalent to non-linear income taxes. Second, the increasing pattern of the dashed black line indicates that, under the US system, childless households always face a positive marginal tax on income. Third, the child allowances paid to mothers with children below 6 imply a negative marginal income tax, as indicated by the decreasing segment of the dash-dotted red line, for earnings below $15,000.[12]

[11] For a single childless person, EITC benefits are phased-in at a rate of 7.65% up to a maximum of $457 in benefits. Families with children benefit from much more generous EITC benefits. For example, for a single parent with two children, EITC benefits are phased-in at a rate of 40% up to a maximum of $5,036 in benefits. See Appendix B for more details

[12] While we focus on the federal income tax, some states also impose state income taxes with rates ranging from 0% to 11%. Low income parents would still benefit from a negative marginal tax rate even if we were to take into account the highest marginal tax rate of 11% (Tax Policy Center, 2010c).

3 Model

From the richness of the US child related transfer and subsidy program, a few normative questions emerge naturally. Is it economically sensible to pay a positive child care subsidy to working mothers? Can the same margin be accounted for with properly designed taxes and transfers on labor income? Should the child care subsidy rate depend on earned income? If yes, should marginal taxes for working mothers be adjusted relative to those levied on childless households? And should the child care subsidy rate depend on total child care cost? In particular, should there be a cap above which the subsidy rate is zero?

In order to address these questions, a flexible economic model is needed, where rich patterns of income taxes and child care subsidies can be studied. The framework presented in this Section, introduces the possibility of engaging in household provided child care in an optimal (non-linear) tax and transfer problem à la Mirrlees in a centralized economy. This relatively simple model captures, we believe, some of the key trade-offs faced by working mothers. We address the optimal design of a tax and subsidy scheme that implements the optimum in a decentralized economy in Section 5.

Agents and Technologies Consider an economy with a continuum of agents who are heterogeneous in market productivities z. We consider discrete levels of market productivity, with $z_1 = 0$ being the minimum and $z_N > 0$ the maximum, that is, $z \in Z := \{z_1, \ldots, z_i, \ldots, z_N\}$. Agents of type z_i constitute a fraction $\pi(z_i) > 0$ of the population, with $\sum_{i=1}^{N} \pi(z_i) = 1$. We interpret agents with $z_1 = 0$ as agents who are subject to adverse labor market conditions (the involuntarily unemployed or unlucky), thereby rendering their market productivity zero.

Agents can allocate effort to market work or to household child care activities. An agent who devotes $l \geq 0$ units of effort on the market produces $y = zl$ of consumption goods. Each agent has child care needs that are normalized to 1 unit of effort, and devote effort level $h \geq 0$ towards them. The remaining amount of child care is covered by purchasing child care from the formal child care market at cost ω per unit.[13] We assume that $z_N > \omega > 0$.[14]

[13] We interpret child care needs as the amount of child care time that can be substituted for paid care during a normal working week. In other words, while $h = 0$ implies that full time formal child care is employed, it does not necessarily imply that mothers never look after their children. For example, mothers could still be taking care of their children during evenings after work.

[14] Whenever either one of the inequalities is not satisfied, our framework specifies into a standard Mirrlees optimal tax model. First, as it can be seen by analogy to the proof of Proposition 1(iii) below, when $\omega = 0$ then

Agents' utility function is additive in consumption c and effort cost $v(e)$:

$$c - v(e),$$

where $e = l + h$ is total effort and c represents household consumption net of formal child care cost $f := \omega (1 - h)$.

Assumption 1 *We assume that the cost function is strictly increasing and strictly convex: $v'(e) > 0$ and $v''(e) > 0$ for all e. In addition, assume that $v'(0) = 0$.*

Laissez-Faire Equilibrium Suppose that agents face no taxes nor subsidies and there are no insurance markets. They solve

$$\max_{l \geq 0, h \geq 0} zl - \omega (1 - h)^{+} - v(l + h),$$

where $(1 - h)^{+} := \max\{0, 1 - h\}$. In the laissez-faire equilibrium, high productivity agents specialize into employment while low productivity agents provide household child care. If $z > \omega$, they optimally choose $h = 0$ and $l > 0$. These high productivity agents consume $c = zl - \omega$ and labor supply solves $z = v'(l)$. When agents have $z < \omega$, they all choose $h > 0$. Low productivity agents with employment opportunities ($0 < z < \omega$) may also work after all child care needs have been taken care of, that is, if $h = 1$. Since, household child care does not depend on labor market productivities, all unemployed agents engage in the same level of household child care and enjoy the same consumption. On the other hand, among employed agents, both earnings and consumption increase in z.

Government and Information Consider a government who aims at distributing resources across agents to maximize welfare. The government does not observe market productivities. The government, however, knows the probability distribution of the different types of agents among the population. The government cannot observe labor supply while it can observe output from the labor market (labor earnings, y), and the total cost of formal child care purchased by each agent (f). Since $f = \omega (1 - h)$, household child care (h) is verifiable (while leisure is

$h(z) = 0$ for all z. In addition, from Proposition 2(a) below, if $z_N \leq \omega$ then all agents will either be pooled into unemployment: $0 < h(z) < 1$ and $y(z) = 0$ for all z, or engage in full-time household child: $h(z) = 1$ and $y(z) \geq 0$ for all z.

not observable). For the purpose of the present application, we endow the government with the amount M of resources to be shared among agents. We interpret M as resources allocated to the group of agents we are interested in (i.e., single mothers with young children), which are obtained from general taxation or other sources that are not studied in this paper. By the revelation principle, we can restrict ourselves to direct mechanisms defined over Z.

Definition 1 *An* allocation *consists of consumption functions $c : Z \rightarrow \mathbb{R}$, market production functions $y : Z \rightarrow \mathbb{R}_+$, and household-provided child care functions $h : Z \rightarrow \mathbb{R}_+$, for all types. Let Ω be the set of such allocations.*

The government also has to satisfy the budget constraint, which can be written as follows:

$$\sum_{i=1}^{N} \pi(z_i)\, c(z_i) + \omega \leq \sum_{i=1}^{N} \pi(z_i)\left[y(z_i) + \omega h(z_i)\right] + M. \tag{1}$$

Modeling the problem as though the government confiscates all production and assigns consumption and child care, is equivalent to imposing a net tax on each agents of type z of $T(z) := y(z) + \omega(1 - h(z)) - c(z)$. Constraint (1) is hence equivalent to $\sum_i \pi(z_i)T(z_i) + M \geq 0$.

The government faces the standard trade-off between redistributing resources and preserving work incentives. In the Laissez-faire allocation, utility increases in z among employed agents and the unemployed get the lowest utility level. Should the government provide too generous redistribution towards low z types, high z types would be tempted to mimic low z types by decreasing effort.

Constrained Efficient Allocation (Second-Best) Since each agent has private information on market productivity, the government faces a set of incentive compatibility constraints. The incentive constraints guarantee the truthful revelation of agents' type z. Agents will only reveal their true type if government policy is such that utility from telling the truth is higher than utility from pretending to be a different type.

Definition 2 *A* reporting strategy *is a mapping $\sigma : Z \rightarrow Z$. By the revelation principle, the planner aims at implementing the* truth-telling strategy, *σ^*, where $\sigma^*(z) = z\ \forall z \in Z$.*

With private information, government allocation has the same domain as above but is based on agents' declarations σ. The definition of an allocation must be re-interpreted accordingly, but still follows Definition 1.

Let

$$V(\sigma|z) := c(\sigma) - v\left(\frac{y(\sigma)}{z} + h(\sigma)\right)$$

be the utility that agent of type z obtains by pretending to be of type σ. The government must guarantee that the agent prefers the truth-telling strategy to any other strategy. Truth-telling requires that for all $z \in Z$,

$$V(z|z) \geq V(\sigma|z) \qquad \forall \sigma \in Z. \tag{2}$$

A key question in the design of an efficient welfare program is how to optimally trade-off redistribution for effort incentives. The objective of the government is to maximize welfare:

$$W(c, y, h; \phi) = \sum_i \pi(z_i)\, \phi(z_i) \left[c(z_i) - v\left(\frac{y(z_i)}{z_i} + h(z_i)\right)\right], \tag{3}$$

where the function $\phi : Z \rightarrow \mathbb{R}_+$ defines the social weighting given by the authorities to the different agents' classes $z \in Z$.

Definition 3 *A* second best allocation *is a solution to the maximization of the objective (3) over* $(c, y, h) \in \Omega$ *subject to the budget constraint (1) and the incentive constraints (2).*

4 The Optimal Allocation

In this Section, we characterize the constrained efficient (second best) allocation. In a standard Mirrlees problem with unidimensional choice of effort, it is customary to use a 'local approach' (i.e., solve the *relaxed problem* that only imposes local incentive compatibility constraints). Under the standard assumption that preferences satisfy the 'single-crossing property of indifference curve maps' (i.e., the marginal rate of substitutions between the choices y and c are monotone in agent's type z), the solution derived from the relaxed problem coincides with the solution to the global problem. In addition, a robust result in the standard optimal taxation model is that one can focus on (local) downwards incentive constraints and hence always obtain downwards distortions, that is, positive labor wedges.

Our model involves a multidimensional choice of effort (work and child care). The monotonicity of marginal rates of substitution between any pair of choices does not suffice the 'single crossing property of indifference curve maps' any more. The most typically adopted approach in the literature on multidimensional choice is to still use a local approach and look for conditions that guarantee that the solution to the relaxed problem deliver a uniformly monotone allocation.[15] Unfortunately, in our framework, uniform monotonicity of the optimal allocations cannot easily be guaranteed a priori. We will hence follow a non-local approach.[16] We look for conditions that guarantee what Matthews and Moore (1987) refer to as *double crossing*. This, in turn, allows us to only focus on downward incentive constraints (see Lemma 1 below). As shown in Lemma 2 in Appendix A, Assumption 2 below guarantees that the utility levels generated by any two allocations, $(\bar{c}, \bar{y}, \bar{h})$ and $(\hat{c}, \hat{y}, \hat{h})$, cross no more than twice in the z space (see Figure 7 in Appendix A).

Assumption 2 *Let $e > 0$. The ratio $\frac{v''(e)}{v'(e)}$ is decreasing in e.*

Standard cost functions such as the quadratic, the constant Frisch elasticity: $v(e) = \frac{1}{\theta}\frac{e^{1+\gamma}}{1+\gamma}$, $\theta, \gamma > 0$, and the exponential cost functions, satisfy this assumption.

An analytical derivation of the constrained efficient allocation also requires an assumption on the social weighting function $\phi(\cdot)$.

Assumption 3 *Let $\mathbf{E}[\phi] := \sum_{i=1}^{N} \pi(z_i)\phi(z_i)$. We have $\phi(z_1) \geq \mathbf{E}[\phi]$; Moreover, for $j \geq 3$, the weight $\phi(z_j)$ is lower than the average social welfare weight: $\phi(z_j) \leq \mathbf{E}[\phi]$.*

Note that Assumption 3 is satisfied by the Utilitarian social welfare function with equal weights $\phi(z_i) \equiv 1$ on all agents. In this case, however, the allocation would display no trade-off between efficiency and redistribution. At the other extreme, the conditions of Assumption 3 are satisfied by the Rawlsian welfare function: $W^R(c, e) := \min_i \{c(z_i) - v(e(z_i))\}$. As we will see below, incentive compatibility implies that $c(z_i) - v(e(z_i))$ increases with i, and hence, the Rawlsian criterium implies $\phi(z_1) > 0$ and $\phi(z_i) = 0$ for $i > 1$. The Rawlsian criterium can be

[15] This is what Matthews and Moore (1987) refer to as 'attribute ordering'. For example, since both the marginal rates of substitution between $(-c)$ and y, and between $(-c)$ and h decrease with z, if y and h were either *both* monotone increasing or both monotone decreasing in z, the allocation would satisfy the single crossing property for the agent's problem and hence local incentive constraints would imply global incentive compatibility (see Lemma 0 in Matthews and Moore (1987)). See also Fudenberg and Tirole (1991), Section 7.3.

[16] Besley and Coate (1995), in Section VII, solve a model similar to ours using a local approach and assuming monotonicity of the marginal rates of substitution. Crucially, they also assume that $\omega = 0$ and $z_1 > 0$. This implies that all agents are optimally required to choose $h = 0$. Their model, hence, reduces to a version of the standard Mirrlees framework where the monotonicity of the marginal rates of substitution implies single crossing of the indifference curve maps.

seen as the limit case for the following class of welfare objectives:

$$\hat{W}(c, e; \rho) := \left(\sum_{i=1}^{N} [c_i - v(e_i)]^\rho \right)^{\frac{1}{\rho}},$$

for $\rho \to -\infty$. Intuitively, for ρ finite but sufficiently low, the implied Pareto weights satisfy Assumption 3. Although it allows for non-monotone ϕ's, Assumption 3 is satisfied whenever the government has a sufficiently strong desire for redistribution at the bottom.[17]

Lemma 1 (Downward IC Approach) *Under Assumptions 1, 2 and 3, any solution to the second best problem where only downward incentive constraints are imposed - that is, when the set of conditions (2) is relaxed to be $\sigma \leq z$ - delivers an optimal allocation. In addition, the 'local' downward incentive constraints can be imposed as equalities. Finally, if the upward incentive constraint is binding for two types $z_j < z_k$, then it is optimal for all agents with type $z_i : z_j \leq z_i \leq z_k$ to receive the same allocation (i.e, bunching).*

Proof. See Appendix A. ∎

Lemma 1 states that the solution from the relaxed second best problem, where the government maximizes the objective (3) subject to the budget constraint (1) and only the downward incentive compatibility constraints in (2), delivers a solution to the original problem.

Given the relaxed problem with downward incentive constraints (DIC) only, we show that the local downward incentive constraints (LDIC) must be satisfied with equality. This crucially relies on the fact that preferences satisfy the double crossing property. Should LDIC between type z_{i+1} and type z_i be slack, then the double crossing property implies that the non-local DIC for preventing type z_{i+1} from mimicking lower types will also be slack. It would therefore be possible to improve welfare at no additional cost and without violating incentives, by redistributing from type z_{i+1} to all other types. Under Assumption 3, such redistribution will weakly improve welfare. It is then easy to show that when the LDIC bind, the upward incentive constraints (UIC) will also be satisfied.

From now onwards, we indicate the allocation obtained using Lemma 1 as 'the optimal allocation', and we denote it by adding an asterisk as superscript.

[17] The requirement that $\phi(z_1) \geq \mathbf{E}[\phi]$ guarantees a well-defined problem and it can be replaced by a participation constraint. ϕ is typically assumed to be non-increasing so that $\phi(z_1) \geq \mathbf{E}[\phi]$ will be automatically satisfied.

Proposition 1 (Minimal Properties) *Under Assumptions 1, 2 and 3, we have:*

(a) The 'net surplus' $y^*(z) + \omega h^*(z) - c^*(z)$ is non-decreasing in z;

(b) Utility of agents in equilibrium $V^*(z|z)$ is non-decreasing in z, and strictly increasing between any two levels $z_{i+1} > z_i$ when $y^*(z_i) > 0$.

(c) For all z, $h^*(z) \leq 1$.

Proof. See Appendix A. ■

Points (a) and (b) in Proposition 1 summarize a general principle. Obtaining a larger net surplus from high types is the sole reason why the government is ready to trade-off redistribution and screen agents instead of pooling them. The last part of Proposition 1 states that providing household child care beyond child care needs would be costly in terms of effort without yielding any additional returns. In particular, this implies that providing $h > 1$ does not help satisfy the incentive constraints. This is because consumption is a superior instrument to achieve separation between types.

Proposition 2 (Characterization) *Under Assumptions 1, 2, and 3, we have:*

(a) Unemployment: Recall that $z_1 = 0$. We have $y^*(z_1) = 0$ and $h^*(z_1) > 0$, where

$$1 - \frac{1}{\omega} v'\left(h^*(z_1)\right) \geq 0, \tag{4}$$

with equality whenever $v'(1) \geq \omega$. If $v'(1) \leq \omega$, then $h^*(z_1) = 1$. In addition, for all z such that $y^*(z) = 0$, type z gets the same allocation as type z_1.

(b) Low productivity: Let $z \leq \omega$. We have $h^*(z) > 0$, and if $y^*(z) > 0$, then $h^*(z) = 1$.

(c) Segmentation: If $y^*(z) > 0$, then $y^*(z') > 0$ for all $z' > z$.

(d) Monotonicity: Let $z' > z$ for which we have no bunching. If $h^*(z') \leq h^*(z)$, then $y^*(z') > y^*(z)$; and if $y^*(z') \leq y^*(z)$, then $h^*(z') > h^*(z)$.

(e) Wedges for the employed: Let z_i be such that $y^*(z_i) > 0$. Then labor wedges are non-negative:

$$1 - \frac{1}{z_i} v'\left(e^*(z_i)\right) \geq 0; \tag{5}$$

If, in addition, $h^*(z_i) > 0$, then the child care wedges are also non-negative:

$$1 - \frac{1}{\omega} v'\left(e^*(z_i)\right) \geq 0. \tag{6}$$

Both wedges are strictly positive whenever $\phi(z_{i+1}) < \mathbf{E}[\phi]$.

For $i = N$, the labor wedge is zero and $h^*(z_N) = 0$.

Proof. See Appendix A. ∎

The intuition for result (a) is simple. When $y(z) = 0$, market productivity does not matter anymore so that all agents receive the same allocation, that is, we have pooling among the unemployed. Result (b) states that low market productivity types may provide positive labor supply only when all child care needs have been met. Statement (c) delivers a minimal monotonicity condition: if an agent is employed, then more productive agents will also be employed. Statement (d) concludes the monotonicity properties of the allocation. Wedges in (e) are direct consequences of the fact that, in our model, only downward incentive constraints matter.

5 The Shape of Efficient Child Care Subsidies

As we have seen in Section 2 (e.g., Figure 1), the existing child care subsidy scheme is rather complex. First, it involves only a partial coverage of formal child care costs. Second, the coverage is nonlinear: the subsidy has a formal child care expenditure cap above which the subsidy rate is reduced to zero. Third, the subsidy rate decreases with labor income. We are interested in understanding whether such features follow from optimality principles.

In this Section, we propose a tax/subsidy scheme that implements the constrained efficient allocation in a decentralized economy. We note that while Assumptions 2 and 3 are sufficient conditions that allow us to analytically characterize the optimal allocations, we do not need to impose those assumptions for our implementation exercise. In other words, our proposed implementation is more general and prevents both upward and downward deviations in the global problem.

5.1 Child Care Wedges and 'Joint Deviations'

As indicated in (16), point (e) of Proposition 2, it is optimal to have the marginal rate of substitution between consumption and child care lower than the return to child care (in consumption terms) for certain agents. Such discrepancies are known as *wedges* in public finance. If agents could freely choose child care (that is not necessarily socially optimal), wedges will be eliminated. A typical way to preserve wedge is to use a tax policy. In our case, a positive subsidy on child care would reduce the privately perceived return to child care and generate a wedge qualitatively similar to that described above. In our framework, however, the relationship between the wedge and the optimal subsidy on child care is not so straightforward. Instead, we show that the optimal subsidy must be kinked as a function of the level of formal child care cost, very much in line with the qualitative features of the existing scheme in the US. An agent whose expenditure on formal child care is lower than the kink point faces a subsidy while it is optimal to set the subsidy to zero (or even to perhaps impose a positive tax) for formal child care cost above the kink-point.

The reason for why the connection between wedges and taxes breaks down in our framework is as follows. The wedge (16) is calculated by figuring out the shadow return to child care of an agent who produces the socially optimal quantities as a function of her skills. Setting the subsidy equal to this wedge eliminates the agent's desire to provide suboptimal child care *when she produces the socially optimal quantities associated with her z type*. However, in a market economy with taxes, an agent might find it optimal to adopt a *joint deviation* of producing a different amount and adjusting the level of child care provided. An optimal tax and subsidy schedule has to be designed so as to deter such joint deviations.

In order to more formally grasp the economic forces shaping child care subsidies in our framework, consider the 'local' wedge as in (16):

$$WE(z_i|z_i) := 1 - \frac{1}{\omega} v' \left(\frac{y^*(z_i)}{z_i} + h^*(z_i) \right).$$

Let $h^*(z_i) < 1$. Suppose that the government is able to induce agent z_i to produce $y^*(z_i)$. $WE(z_i|z_i) \geq 0$, hence, represents a necessary condition for the agent to choose $h^*(z_i)$.

Setting marginal income tax rates equal to the labor wedges (15) and marginal child care subsidy rates equal to the child care wedges $WE(z_i|z_i)$, however, will not be enough to implement the constrained optimum. This is because those who tell the truth about their type would not be the only ones who would want to increase h. In fact, higher types who declare to be of a type $\sigma = z_i$ will have even greater incentives to overprovide h (while also engaging in suboptimal market work). In particular, consider agent z_{i+1} declaring to be of type z_i. The 'joint deviation wedge' for this agent is given by:

$$WE(z_i|z_{i+1}) := 1 - \frac{1}{\omega} v' \left(\frac{y^*(z_i)}{z_{i+1}} + h^*(z_i) \right).$$

Clearly $WE(z_i|z_{i+1}) > WE(z_i|z_i)$, that is, agents of type $z_{i+1} > z_i$ face a joint deviation child care wedge that is larger than the child care wedge for a true-telling agent of type z_i. In other words, if we were to set the child care subsidy rate to $WE(z_i|z_i)$, then agent z_{i+1} pretending to be of type z_i and producing the recommended level of income $y^*(z_i)$ for this declaration, finds it optimal to increase h beyond $h^*(z_i)$. This is problematic since, as shown in Lemma 1, the LDIC is binding at the optimal allocation. This implies that, whenever the child care subsidy rate is set equal to $WE(z_i|z_i)$, agent z_{i+1} finds it strictly more advantageous to declare $\sigma = z_i$, produce $y^*(z_i)$ and choose $h > h^*(z_i)$ compared to declaring the truth (and choosing the recommended values $(y^*(z_{i+1}), h^*(z_{i+1}))$ for his type). These complications are even stronger when non-local DIC are binding, a non-pathological feature of the optimal allocation in our multidimensional choice setting. For the purpose of implementing a second best allocation, it is therefore important to consider the possibility of joint deviations in declaring a different type σ and engaging in a non-optimal level of h.

A Graphical Representation of the Optimal Child Care Subsidy Schedule The rational behind the efficient subsidy scheme can be seen graphically as follows. Recall that $V^*(\sigma|z)$ is the value for agent z of declaring σ according to the constrained efficient allocation:

$$V^*(\sigma|z) := c^*(\sigma) - v\left(\frac{y^*(\sigma)}{z} + h^*(\sigma)\right),$$

where $(c^*(\sigma), y^*(\sigma), h^*(\sigma))$ are the constrained optimal allocations associated with type σ. Second best optimal net taxes are given by:

$$T^*(\sigma) = y^*(\sigma) - c^*(\sigma) - \omega(1 - h^*(\sigma)).$$

Suppose now that agents can privately choose which type to declare, $\sigma \in Z$, as well as household provided child care. Taking the second best optimal $y^*(\sigma)$ and $T^*(\sigma)$ as given, an agent z therefore chooses σ and h so as to maximize her private utility:

$$\max_{\sigma,h} \quad \underbrace{y^*(\sigma) - T^*(\sigma) - \omega(1-h)^+}_{\text{c}} - v\left(\frac{y^*(\sigma)}{z} + h\right). \tag{7}$$

If each agent who reports σ engages in the constrained efficient level of household child care associated with type σ (i.e., $h = h^* (\sigma)$), then incentive compatibility would imply that all agents would reveal their true type. A necessary condition for this to happen is that the agent faces a subsidy that solves her first order condition with respect to household child care at $h^* (\sigma)$. We would thus require a subsidy rate equal to the joint deviation child care wedge at $h = h^* (\sigma)$. Let $s (\sigma|z)$ be such a rate:

$$s (\sigma|z) = WE (\sigma|z) := 1 - \frac{1}{\omega} v' \left(\frac{y^* (\sigma)}{z} + h^* (\sigma) \right).$$

Hence, we have:

$$(1 - s (\sigma|z)) \omega - v' \left(\frac{y^* (\sigma)}{z} + h^* (\sigma) \right) = 0.$$

We illustrate the private maximization problem (7) of an agent of type z declaring to be of type σ in Panel (a) of Figure 2. In the absence of child care subsidies, the slope of the budget constraint, $c = y^* (\sigma) - T^* (\sigma) - \omega (1 - h)$, is equal to the cost of formal child care ω. Agent z declaring σ engages in household child care $h (\sigma|z) \in (0, 1)$ given by the tangency point between the agent's indifference curve and the agent's budget constraint at point A. To implement the constrained optimum, we need to induce any agent who delcares σ to choose the constrained optimum level of household child care, $h^* (\sigma)$. A child care subsidy rate set equal to the joint deviation wedge of the agent at $h^* (\sigma)$ ensures that the slope of the budget constraint becomes $(1 - s (\sigma|z)) \omega$. Agent z declaring σ will therefore choose $h^* (\sigma)$ at point B.

This hypothetical subsidy scheme is, however, infeasible since the subsidy rates are dependent on the true type z of the agent, which is nonobservable. We therefore need to design a subsidy scheme that does not rely on observing z. Suppose that, as in Figure 2(a), in the absence of child care subsidies, an agent z reporting σ has incentive to engage in $h > h^* (\sigma)$. Such deviation, would be discouraged by setting the subsidy rate equal to the joint deviation wedge of highest type z_N:

$$WE(\sigma|z_N) = 1 - \frac{1}{\omega} v' \left(\frac{y^* (\sigma)}{z_N} + h^* (\sigma) \right).$$

Since $WE (\sigma|z_N) \geq WE (\sigma|z)$ for all z, no z declaring σ would ever choose h above $h^* (\sigma)$. Symmetrically, setting a subsidy rate equal to $WE(\sigma|z_2)$ guarantees that each agent z reporting σ has an incentive to choose $h \leq h^* (\sigma)$. Such a scheme is illustrated by the solid red lines in Panel (b) of Figure 2. The scheme displays a kink-point at $h^*(\sigma)$. At point B in Figure 2(b), the steeper segment of the kinked budget constraint (in red) is tangent to the indifference curve

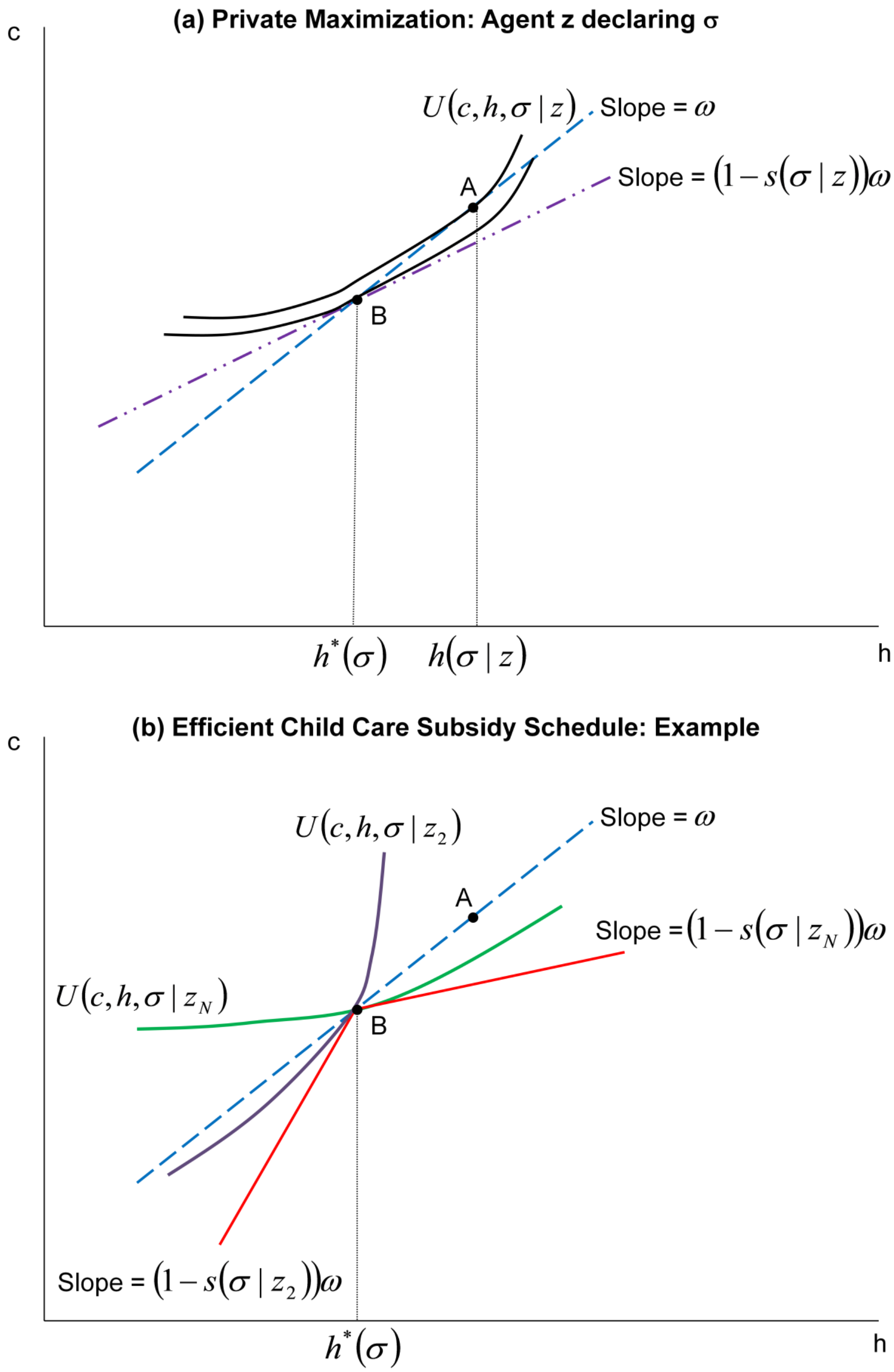

Figure 2: $U(c, h, \sigma|z)$ corresponds to the objective function in (7). *Panel (a):* In the absence of child care subsidies, agent z declaring σ engages in household child care level $h(\sigma|z)$, given by the tangency point between the agent's indifference curve and the agent's budget constraint at point A. A child care subsidy rate set equal to the joint deviation wedge of the agent at $h^*(\sigma)$ would ensure that an agent z declaring σ will choose $h^*(\sigma)$ at point B. *Panel (b):* A subsidy rate that is set equal to the maximum joint deviation wedge $s(\sigma|z_N) = WE(\sigma|z_N)$ when $h \geq h^*(\sigma)$ and to the minimum joint deviation wedge $s(\sigma|z_2) = WE(\sigma|z_2)$ when $h < h^*(\sigma)$, ensures that any agent declaring to be of type σ chooses the optimal level of household child care $h^*(\sigma)$. An example of such a scheme is depicted by the red solid line budget constraint with a kink at $h^*(\sigma)$.

for agent z_2 (in purple) while the flatter segment of the kinked budget constraint is tangent to the indifference curve for agent z_N (in green). Since the indifference curve of any z reporting σ would lie in between the indifference curves associated with z_2 and z_N at the kink point, any agent reporting σ would choose $h^*(\sigma)$. This principle is used in Proposition 3, where we also show that z_2 can be replaced by the productivity level of the highest unemployed type.

5.2 Implementation

We first discuss an implementation that relies on direct mechanism and subsequently map our proposed implementation using a version of the taxation principle.

Recall, for any real number x, we adopt the notation $x^+ := \max\{0, x\}$ and $x^- := \min\{x, 0\}$. Let $Z_0^* := \{z \in Z \mid y^*(z) = 0\}$ the set of types pooled into unemployment, and $\bar{z}_0 := \max Z_0^*$ the highest type in this set.

Proposition 3 *Let $f^*(\sigma) := \omega(1 - h^*(\sigma))$ be the optimal formal child care cost associated with the optimal $h^*(\sigma)$. The following subsidy rates and transfers implement the constrained optimum.*
(a) For employed agents, we have:

$$\text{If} \quad \sigma \notin Z_0^*, \quad \text{then} \quad s(\sigma, f) = \begin{cases} \left(1 - \frac{1}{\omega} v'\left(\frac{y^*(\sigma)}{z_N} + h^*(\sigma)\right)\right)^+ & \text{if} \quad f \leq f^*(\sigma); \\ \left(1 - \frac{1}{\omega} v'\left(\frac{y^*(\sigma)}{\bar{z}_0} + h^*(\sigma)\right)\right)^- & \text{if} \quad f > f^*(\sigma). \end{cases}$$

(b) For unemployed agents, the subsidy rate is zero: If $\sigma \in Z_0^*$, then $s(\sigma, f) = 0 \; \forall f$.
(c) For all $\sigma \in Z$, the optimal transfer scheme is set as follows:

$$T(\sigma) = y^*(\sigma) - c^*(\sigma) - f^*(\sigma) + s(\sigma, f^*(\sigma)) f^*(\sigma);$$

where $c^*(\cdot)$ and $y^*(\cdot)$ are the consumption and income functions of the second best allocation.

Proof. See Appendix A. ■

The identification of the type $\bar{z}_0$ permits one to minimize the subsidy rates in the second segment of the subsidy schedule while analytically guaranteeing the implementation of the second best. The operators x^+ and x^- have a similar aim. They imply that the child care subsidy rate is set to zero whenever such zero rate is 'analytical sufficient' to implement the second best.

As described above, child care subsidies in Statement (a) ensure that each agent declaring σ chooses the optimal level of formal child care cost associated with σ, $f^*(\sigma)$, no matter what her true type is. As can be seen from Statement (c), income taxes are then adjusted to yield the same consumption to agents as in the constrained optimum: $c^*(\sigma) = y^*(\sigma) - T(\sigma) - (1 - s(\sigma, f^*(\sigma))f^*(\sigma)$. Since agents earn the same and receive the same consumption levels as in the second best optimum, such allocations are incentive compatible and also satisfy the government budget constraint.

We note that for employed agents, if $f^*(\sigma) = \omega$, then only the subsidy rates associated with the first segment $f \leq f^*(\sigma)$ are relevant. Similarly, if $f^*(\sigma) = 0$, then only the subsidy rates associated with the second segment $f > f^*(\sigma)$ are relevant. We will see in our simulation exercises in the next Section, that most employed mothers choose $f^*(\sigma) = \omega$, i.e., $h^*(\sigma) = 0$.

Statement (b) deals with child care subsidies offered to the unemployed. Since market productivities are irrelevant for the unemployed, they are all the same and there are no incentives problem among them. There is therefore no need to subsidize child care of the unemployed.

The implementation is straightforward in the sense that we do not need to compute who deviates where and by how much. In other words, we do not need to compute all the joint deviation wedges. The child care subsidies are conditional on formal child care cost being verifiable.

The optimal subsidy rates and transfers schedule englobes features that match the qualitative features of the US system, that is, a cap on formal child care costs and subsidy rates that decrease with earnings for formal child care costs below the cap. We propose such a scheme using a variation of the taxation principle below.

To be able to describe the subsidy rates and transfer scheme as only a function of income, we need an additional monotonicity assumption. We abuse in notation and indicate by $f(y)$ the formal child care level associated with income y. For all values of y such that there is a σ_y: $y = y^*(\sigma_y)$, we associate $f(y) = f^*(\sigma_y)$. Unfortunately, such mapping does not deliver a well-defined function whenever the optimal allocation associates multiple values of f to one y. A natural assumption that guarantees a well-defined function $f(\cdot)$ is monotonicity.

Assumption 4 *Let $\mathcal{Y} = \{y \in \mathbb{R}_+ | \exists z \in Z \ : \ y = y^*(z)\}$ be the set of equilibrium income values, and for all $y \in \mathcal{Y}$ define $f(y) := f^*(\sigma_y)$. Assume that $f(\cdot)$ is non-decreasing in $\mathcal{Y}$.*

As we will see in the numerical section (Section 6.2), in all our simulations, f turns out to be non-decreasing in y. Under Assumption 4, we can extend the domain of $f(\cdot)$ to $\mathbb{R}_+$ by setting for $y \geq 0, y \notin \mathcal{Y}$, $f(y) = f(m(y))$ where $m(y) := \max\{\hat{y} \in \mathcal{Y} | \hat{y} \leq y\}$. The consumption function is analogously constructed: $c(y) = c(m(y))$, where for all $y \in \mathcal{Y}$, $c(y) = c^*(\sigma_y)$.

Proposition 4 *Under Assumption 4, there is a $\overline{T} \in \mathbb{R}$ such that the following subsidy rates and transfers implement the constrained optimum.*
(a) For employed agents (who earn $y > 0$), we have:

$$s(y, f) = \begin{cases} \left(1 - \frac{1}{\omega} v'\left(\frac{y}{z_N} + 1 - \frac{f(y)}{\omega}\right)\right)^+ & if \quad f \leq f(y); \\ \left(1 - \frac{1}{\omega} v'\left(\frac{y}{\overline{z}_0} + 1 - \frac{f(y)}{\omega}\right)\right)^- & if \quad f > f(y); \end{cases}$$

$$\text{If } y \in \mathcal{Y}, \text{ then } T(y) = y - c(y) - f(y) + s(y, f(y)) f(y); \text{ Otherwise, } T(y) = \overline{T}.$$

(b) For unemployed agents (with $y = 0$), the second best allocation is implemented by having:

$$s(0, f) \equiv 0, \qquad \text{and} \qquad T(0) = -c(0) - f(0).$$

Proof. See Appendix A. ∎

Note that when when $f = \omega$, the child care subsidies for the employed with $f < f(y)$ follow a *sliding scale* pattern, that is, they decrease with labor earnings. We will see in the next Section, that child care subsidies follow a sliding scale in our simulation exercises.

In our implementation, labor wedges and marginal taxes on income do not coincide. Using the private first order condition of the agent with respect to y, evaluated at the agent's optimal formal child care cost choice $f(y)$, we obtain:

$$T'(y) = 1 - \frac{1}{z} v'\left(\frac{y}{z} + 1 - \frac{f(y)}{\omega}\right) + s'_y(y, f(y)) f(y). \tag{8}$$

When $s'_y(y, f) \leq 0$, since $f(y) \geq 0$, our implementation implies a marginal income tax that is no greater than the labor wedge.[18] This observation might contribute to the debate over the optimality of imposing a negative marginal income tax rate on low income earners. The debate has focused on the possibility of having *negative labor wedges* whenever there is a strong desire to redistribute towards low skilled individuals (Choné and Laroque, 2011; Saez, 2002a). As we

[18]Note that the function $s(y, f)$ is not differentiable in y at $f = f(y)$. For all practical purposes, however, we can focus on the initial segment of the subsidy rate schedule, where cost of formal child care is below $f(y)$.

saw in Section 2, only working mothers face negative marginal taxes at low income in the US system. At the same time, they also face child care subsidies that decrease with earned income. Equation (8) indicates that when child care subsidies follows a sliding scale, optimal negative marginal taxes can be compatible with positive to labor wedges.

Child Allowances Finally, in order to implement the second best allocation in a way that is compatible with the current US tax schedule, we need to specify optimal income dependent child allowances. Let $T^a(y)$ denote net taxes faced by a childless individual earning y in the actual US tax and benefit system (the corresponding schedule is the dashed black line in Panel (c) of Figure 1). As described in Section 2, the existing child allowances $A^a(\cdot)$ include child related federal income tax exemptions and EITC if employed, and TANF benefits if unemployed (this schedule corresponds to the solid blue line in Panel (c) of Figure 1).

We take the general income tax scheme for the childless $T^a(y)$ as given and keep it fixed. *The optimal child care subsidy rates $s(y, f)$ from Proposition 4, together with the optimal child allowances, $A(y)$, implement the constrained optimum,* where child allowances are defined as:

$$A(y) := T^a(y) - T(y),$$

and $T(y)$ are the total optimal transfers from Proposition 4. It is indeed straightforward to see that since child care subsidy rates are the same as in Proposition 4, agents would engage the optimal level of formal child care. In addition, since $T(y) = T^a(y) - A(y)$, Proposition 4 implies that consumption and utility would also be the same as under the second best.

6 Quantitative Analysis

In this section, we present a quantitative analysis based on our framework. We focus on single mothers with at least one child aged below 6 and calibrate our model to match features of the US labor market. We then simulate our optimal policy results and quantify the optimal child care subsidies and child allowances.

6.1 Calibration

Recall the welfare function $\hat{W}(\cdot;\rho)$ discussed in Section 4. A value $\rho = 1$ indicates the lowest desire for redistribution; the laissez-faire allocation with zero taxes and no subsidies solves the government problem for this case. For our benchmark case, we consider a moderate desire for redistribution by assuming a logarithmic welfare function ($\rho = 0$):

$$\sum_{z \in Z} \pi(z) \ \ln\left(c(z) - \frac{1}{\theta}\frac{e(z)^{1+\gamma}}{1+\gamma}\right),$$

where $\frac{1}{\gamma}$ represents the wage elasticity of labor supply.

By taking the derivative with respect to $c(z)$, we recover the (here endogenous) social welfare weights:

$$\phi^*(z) := \frac{1}{V^*(z|z)} = \frac{1}{c^*(z) - \frac{1}{\theta}\frac{e^*(z)^{1+\gamma}}{1+\gamma}}.$$

Although the sufficient condition stated in Assumption 3 is possibly not satisfied by the social weighting function ϕ^*, a simple Corollary of Lemma 1 suggests the following algorithm. Compute the optimal allocation of the relaxed problem with only (local and non-local) downward constraints. If the LDIC are satisfied with equality, all upward constraints are also satisfied and all properties of Lemma 1 hold. In Appendix B.4, we describe the numerical algorithm.

As a second criterium, we will also consider the Rawlsian case where the planner aims at maximizing the welfare of the unemployed. As discussed in Section 4, this objective satisfies Assumption 3 and hence, no ex-post verification is needed in this case.

We have the following parameters to calibrate: the preference parameters γ and θ, the probability of facing adverse labor market conditions $\pi(0)$, the distribution of market productivities $\pi(z)$ when $z > 0$, the child care needs of one unit which corresponds to choosing a normalization for effort e, the cost of formal child care ω, and the amount of net resources allocated to single mothers M. Table 1 summarizes the parameter values and relevant moments used for calibration.

Wage Elasticity From our utility specification, the labor supply elasticity is given by $\frac{1}{\gamma}$. Following the literature on wage elasticity among women (see Heckman and Macurdy, 1980 and Blundell, Meghir, and Neves, 1993), we set $\gamma = 1$ corresponding to an elasticity of 1. We also conduct sensitivity analysis by considering a more conservative elasticity of 0.5, corresponding to preference parameter $\gamma = 2$ (Chetty et al., 2011).

Table 1: **Parametrization**

Parameter	Value	Moments to match	Source
γ	1	Wage elasticity of labor supply 1	Heckman and Macurdy (1980)
			Blundell, Meghir, and Neves (1993)
θ	See Table 3	Average hours of work	CPS 2010
$\pi(0)$	11%	Proportion involuntarily unemployed	CPS 2010
$\pi(z)$	See Figure 3	Empirical distribution of wages	CPS 2010
$e = 1$	24 hours	Hours of non family child care per week	Rosenbaum and Ruhm (2007)
			Laughlin (2010)
ω	\$5.10	Average cost of child care per hour	Child Care Aware of America (2012)
M	See Table 3	Net transfers to single mothers	Federal and SS Tax, EITC
			DCTC, CCDF, TANF

Table 2: **Summary Statistics**

Variable	Mean	s.d.	Variable	Mean	s.d.
Age	28.3	7.29	Proportion in good health	0.89	0.31
High school graduate	0.33	0.47	Proportion working	0.56	0.50
College or university	0.44	0.50	Yearly hours of work (if > 0)	1,519	749
No. of children under 6	1.28	0.55	Wage per hour (if > 0)	14.5	0.49
No. of children under 18	1.95	1.10	Out of the labor force	0.32	0.46
White	0.66	0.47			
Black	0.25	0.43	No. of observations	3,211	

Source: March 2010 CPS data on single women with at least one child aged below 6.

CPS data To calibrate effort cost parameter θ as well as the distribution of market productivities, we make use of March 2010 Current Population Survey (CPS) data. We limit the sample to single mothers aged between 18 and 50, and who have at least one child aged below 6. Table 2 reports summary statistics for our sample of single mothers. On average, our mothers tend to have 1.28 children aged below 6 and 1.95 children aged below 18. 56% of mothers worked and employed mothers on average worked 1,519 hours a year.

Adverse labor market conditions We specify the probability, $\pi(0)$, of people suffering from adverse labor market conditions ($z_1 = 0$) as the proportion of involuntarily unemployed mothers in our CPS sample. The involuntarily unemployed include those who lost their jobs and those who entered or re-entered the labor force but could not get a job. This definition excludes those who voluntarily left their jobs or are out of the labor force. Around 11% of mothers with children under 6 were involuntarily unemployed and represent our mass of people suffering from adverse labor market conditions.

Empirical distribution of productivities In our model, individuals have heterogeneous market productivities when working outside of the house. We interpret market productivity types $z > 0$ as hourly wages when agents are not involuntarily unemployed. A standard approach to calibrating the skills distribution in the literature has been to fit a smooth function (typically log normal with upper Pareto tail) on empirical wages of the employed (Mankiw, Weinzierl, and Yagan, 2009). This approach, however, does not consider the skills distribution among the voluntarily unemployed.

In our framework, agents can be voluntarily unemployed if their wage is lower than their reservation wage which, in our model, depends on formal child care cost as well as the actual US tax and benefit system. In particular, we care about the *potential* wage distribution of mothers who would have worked if they did not have child care needs or faced a more generous child care subsidy scheme. This is because while it may be privately optimal for them not to work given the current real world situation, it may be efficient for some of them to work in the optimal program. We therefore impute the potential wage distribution of the voluntarily unemployed using two step selection correction methods à la Heckman. The empirical strategy and wage regression coefficients are reported in Appendix B.

Figure 3 illustrates the wage distribution of mothers conditional on not being unlucky, $\pi(z)$ for $z > 0$. The wage distribution of the employed is based on their actual hourly wages, which are computed as yearly gross earnings divided by total hours of work in one year. The potential wage distribution of voluntarily unemployed mothers have been imputed. In the calibration exercise, we discretize the wage space into 50 wage centiles ranging between \$2.40 to \$32.21 such that we have 2% of mothers within each bin.

Child care needs Data from the Survey of Income and Program Participation (SIPP) indicates that on average, preschool age children with employed mothers spent 6 hours per week in the care of fathers, 10 hours in grandparent care, and 15 hours in an arranged care facility (Rosenbaum and Ruhm, 2007; Laughlin, 2010). We interpret child care needs as the amount of

Wage Distribution

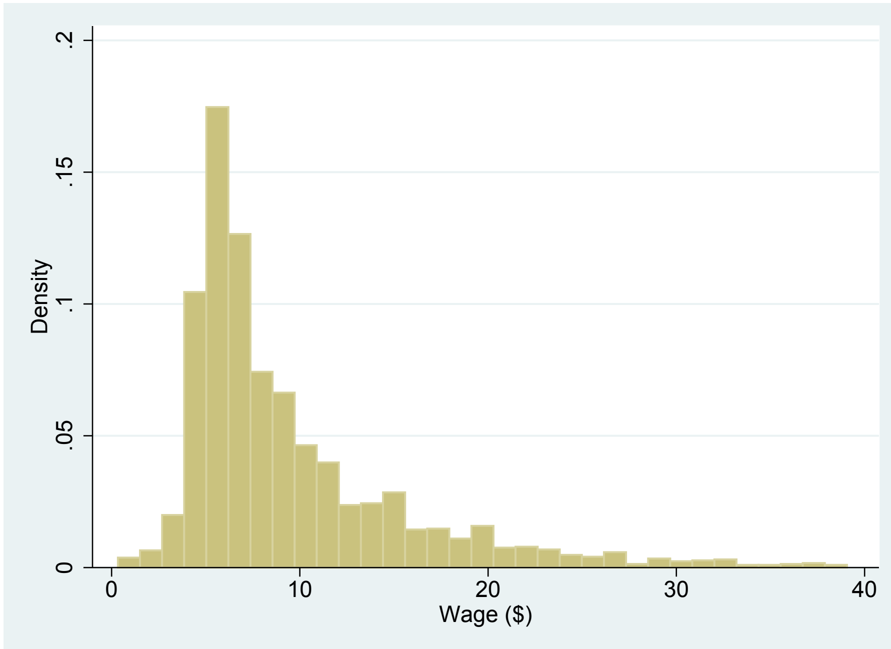

Figure 3: *Source*: March 2010 Current Population Survey data on women with at least one child aged below 6. Wages for non working mothers are imputed using Heckman selection correction methods.

child care time that can be substituted for paid care. Given a normal working week of 40 hours and family provided care of 16 hours per week, mothers need to make alternative child care arrangements for the remaining 24 hours per week. We therefore calibrate our model such that one unit of effort is equal to 24 hours per week. We also perform sensitivity analysis where we use a normalization of 34 hours corresponding to a 50 hour working week minus family provided care of 16 hours per week.

Child care cost To calibrate average hourly cost of formal child care ω, we use the 2010 US average cost which ranged between $6,380 for a four year old in family care homes and $9,520 for an infant in child care centres (Child Care Aware of America, 2012).[19] Assuming that full time child care corresponds to 50 weeks of 40 hours each, we get an average hourly cost of $3.98 for one child. Since single mothers in our sample have on average 1.28 children under age 6, we set $\omega = \$5.10$. As sensitivity check, we also consider a higher $\omega = \$6.40$, corresponding to formal child care cost of $10,000 per year for one child in full time day care.

[19] State wide average annual costs for a four year old (infant) in full time centre based care ranged between $3,900 ($4,600) in Mississippi and $11,700 ($15,000) in Massachusetts.

Calibration of effort cost We calibrate the parameter θ such that, given the 2010 US tax and benefit system and the selection corrected empirical distribution of wages, the average hours of work predicted from our model match the average hours of work observed in the CPS sample. The private problem of an agent of with effort cost parameter θ is given by

$$\max_{\{c,l,h\}} \quad c - \frac{1}{\theta}\frac{e^{1+\gamma}}{1+\gamma} \tag{9}$$

s.t.

$$c = y - T^E(y, f) - f,$$

where effort $e = l + h$, earnings $y = zl$, and formal child care cost $f = \omega(1-h)$. $T^E(y, f)$ are actual net taxes faced by a working mother: Federal and SS Taxes, EITC, DCTC and CCDF benefits.[20] We elaborate on this stage of the calibration process in Appendix B. The calibrated θ for different child care needs, child care cost and labor supply elasticity are reported in Table 3.

Calibration of net transfers Since our analysis is focused on single mothers with children aged below 6 and the US welfare system tends to be generous towards them, we also need to calibrate the amount of net transfers, M, already allocated to them in the US budget. In other words, we take as given the current generosity of the US towards single mothers. M is therefore a weighted sum of unemployment benefits, and of EITC, DCTC and CCDF benefits net of Federal and SS taxes when the mother is employed. The weights are given by the probability distribution of mothers across z types. We interpret unemployment benefits as inclusive of the TANF and set the benefits such that, given the US tax and benefit system, the proportion of working mothers predicted by our model matches the proportion of working mothers (56%) in our CPS sample. We elaborate on this stage of the calibration process in Appendix B. The calibrated values of M for different child care needs, child care cost and labor supply elasticity are reported in Table 3.

Model validity We have made several assumptions while calibrating our model. The main assumptions include the use of the wage distribution to model the distribution of market productivity and our model implication that people with employment opportunities and wages

[20] Recall that although the US child care subsidy rates do not directly vary with the cost of formal child care, there is an upper cap on subsidies that can be received on child care cost under the DCTC.

Table 3: **Calibrated Effort Cost Parameter and Net Transfers**

	Baseline	$1e = 34$	$\omega = 6.4$	$\gamma = 2$
θ	1.27	0.88	1.27	1.77
M	\$4,617	\$5,015	\$4900	\$4,627

Note: θ is calibrated such that, given the 2010 US tax and benefit system and the distribution of wages, the average hours of work predicted from our model match the average hours of work observed for single mothers with at least one child under 6 in 2010 CPS. Baseline specification with $1e = 24$, $\omega = \$5.1$ and $\gamma = 1$. In sensitivity analysis, we recalibrate θ and M by varying the parameter of reference while keeping the other ones at the baseline level.

above the reservation wage would work. As an external validity check, we compute the employment elasticity implied by our model based on the proportion of women who would leave employment as a result of a 2% increase in the cost of formal child care. According to our calibrated model, the employment elasticity with respect to the cost of child care is -0.83, which lies within the average range of child care price elasticities estimated in the literature.[21]

6.2 Results

We simulate the constrained optimal allocations implied by our model and quantify the optimal child care subsidy rates and transfers that implement the constrained optimum.

Constrained Optimal Allocations Results for the constrained optimum are illustrated in Figure 4. The solid lines illustrate the baseline case with a labor supply elasticity of one $\gamma = 1$, a normalization of one unit of effort set equal to twenty four hours $1e = 24$, and cost of formal child care of $\omega = \$5.10$ per hour. We perform sensitivity analysis, where we recalibrate θ and M by varying, respectively, labor supply elasticity to 0.5 corresponding to $\gamma = 2$, the normalisation to $1e = 34$, and the cost of formal child care to $\omega = \$6.40$, while keeping the other parameters at the baseline level.

In all specifications, earnings (y) and optimal consumption (c) increase with market productivity (z), as can be seen from Panels (a) and (b) of Figure 4. As expected, unemployed mothers are pooled with the same consumption and household child care. Single mothers with low

[21] Employment elasticities with respect to cost of child care for US single mothers with children aged below 6 range from -0.5 (Han and Waldfogel, 2001) to -1.29 (Connelly and Kimmel, 2003).

The Optimal Allocation

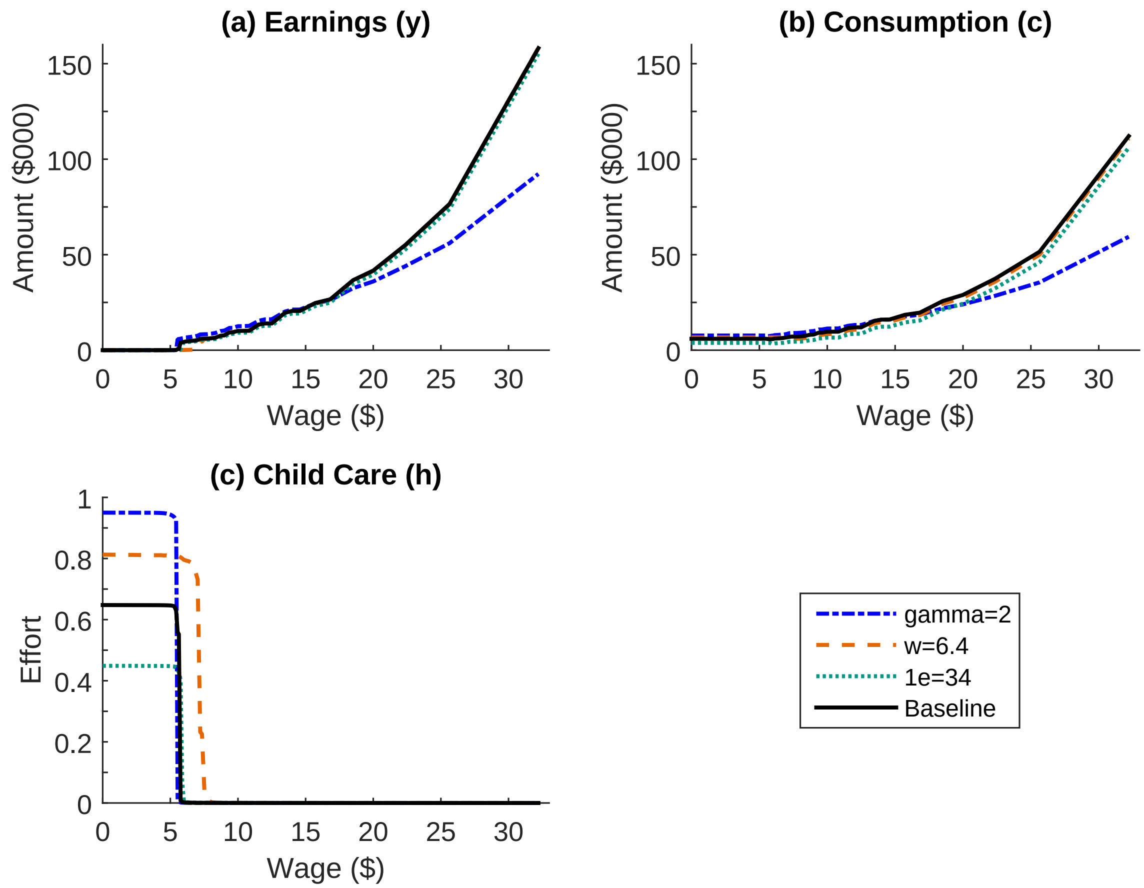

Figure 4: Panel (a) depicts optimal output of employed mothers as measured by yearly earnings. Panel (b) depicts optimal consumption measured in thousands of dollars per year. Panel (c) depicts optimal household provided child care of mothers measured as a fraction of child care needs.

productivity are engaged in higher levels of household child care compared to the benchmark case when the cost of formal child care is high ($\omega = \$6.40$) and when labor supply is more inelastic ($\gamma = 2$), as can be seen from Panel (c).

Working mothers tend not to engage in household child care activities ($h = 0$). Across all specifications and productivity levels, we have only five cases where mothers work on the labor market and also engage in household child care activities: The productivity type $z = 5.88$ in the specification with normalization $1e = 34$, and four productivity types, $z = \$7.21 - \$7.71$, in the specification with high formal child care cost ($\omega = 6.40$).

For the sake of comparison with a different welfare criterion, we illustrate the optimal allocations and utilities implied by both the benchmark case and the Ralwsian social welfare

Comparison with Allocations under 2010 US Tax and Benefit System

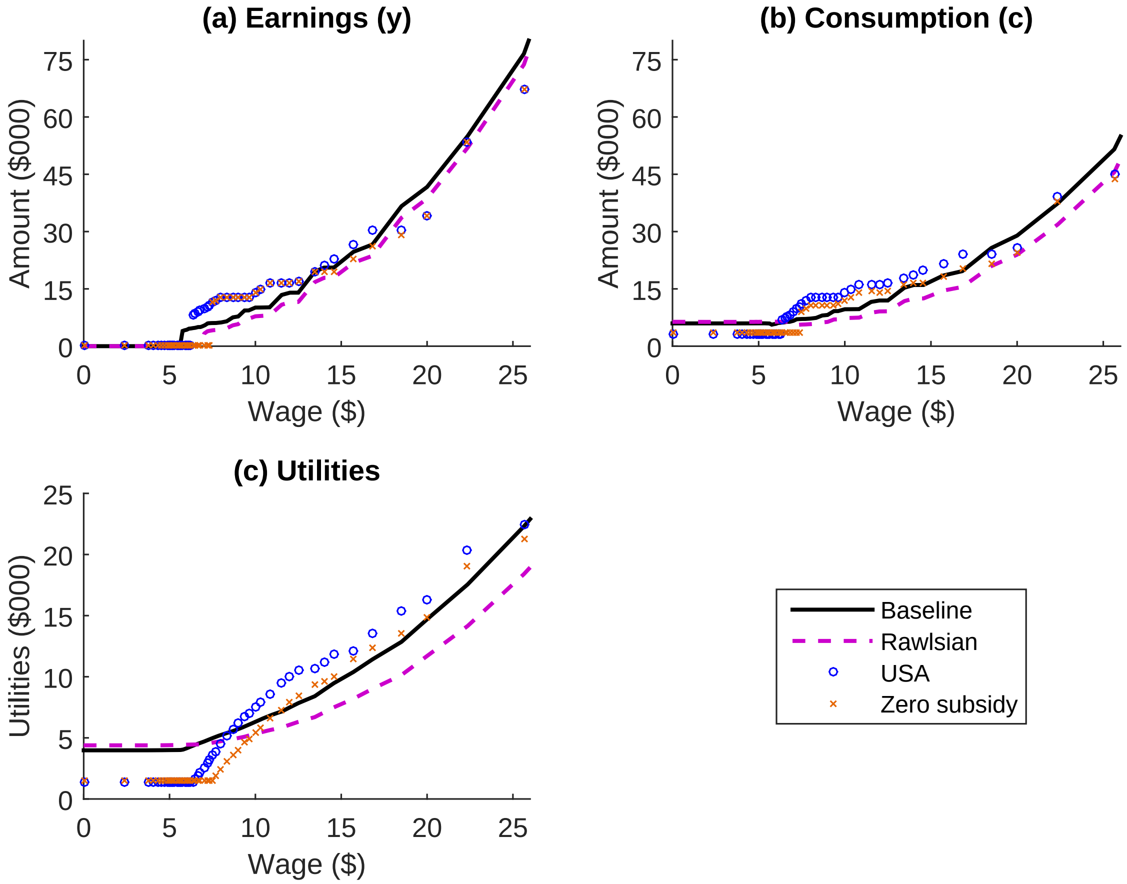

Figure 5: The solid black lines represent features of the optimal allocation according the benchmark case with logarithmic welfare objective, while the fucsia dashed lines represent the same features when the optimal allocation is computed using the Rawlsian criterium. The blue circles represent the simulated allocation obtained from solving the agent problem (9), that is, assuming the agent faces the US tax and benefit scheme for mothers with two children below 13, inclusive of the CCDF and DCTC. The orange crosses represent the allocation obtained from the simulations assuming that the agent faces the US tax and benefit scheme but with child care subsidy rates set to zero (i.e., inclusive of federal and SS taxes, EITC and TANF, but without CCDF or DCTC). Panel (a) depicts yearly earnings, Panel (b) shows the level of consumption, while Panel (c) depicts the implied utilities for each type. The graphs are truncated at wage $26 for emphasis (we do not show the highest wage bin of $32.21).

function in Figure 5. As expected, the unemployed get higher utility under the Rawlsian compared to logarithmic welfare criterion. Under the Rawlsian criterium, the employed work and consume less and get lower utility compared to the benchmark case.

In the same figure, the blue circles represent the imputed allocation and utilities implied by the actual US tax and benefit system under the baseline specification. More precisely, to get the allocations implied by the US tax and benefit system, we simulate the choice of single mothers taking into account federal and SS taxes, EITC, TANF, DCTC and CCDF as per the agent's problem (9). We also report the imputed allocation and utilities from the simulations assuming that child care subsidies are zero (i.e., inclusive of federal and SS taxes, EITC and TANF, but without CCDF or DCTC), as illustrated by the orange crosses.[22]

The disincentive effects of not providing any child care subsidy on labor supply can be seen clearly from Panel (a) of Figure 5, where a higher proportion of mothers do not work. *In the optimal scheme, an even higher proportion of mothers work relative to the US system.* The optimal scheme also provides higher consumption and higher utility to low productivity mothers as can be seen from Panels (b) and (c). The incentive issues at hand suggest that child care subsidies may encourage labor supply, especially among low productivity types and at the extensive margin of participation.

Earnings Related Taxes and Child Care Subsidies Table 4 reports the optimal child care subsidy rates for employed mothers according to the implementation proposed in Proposition 4. We report the subsidy rates associated with the first segment of the schedule, when $f \leq f(y)$. As noted above, across all specifications, we have only 5 cases with $y > 0$ and $0 < f(y) < \omega$. The optimal subsidy rate associated with the second part of the kink when $f > f(y)$ was 0 (zero) in all 5 cases; this shape is representable by a 'cap' such as the existing one in the US. For all other cases, $f(y) = \omega$ so that only the rate in the first segment is relevant.

As can be seen from the third column of the Table, the optimal subsidy rates for the benchmark specification start from 80% and decrease towards zero for earnings of $30,000 or above. *The optimal subsidy rates decrease more steeply that the US ones* (DCTC and CCDF), reported in the second column. This qualitative feature is robust across welfare criteria and parametric specifications, at it can be seen from the remaining columns.

[22]In order to have a meaningful comparison, all reported specifications have the same value of net transfers $M = \$4,617$. The government budget 'net cost savings' from not providing child care subsidies were reallocated equally across all types as consumption in the specification without child care subsidies.

Table 4: **Child Care Subsidy Rates**

Earnings	US	Baseline	Rawls	$1e = 34$	$\omega = 6.4$	$\gamma = 2$
\$5,000	0.70	0.80	0.80	0.80	0.84	1.00
\$10,000	0.70	0.60	0.60	0.59	0.68	0.93
\$15,000	0.70	0.40	0.40	0.39	0.52	0.83
\$20,000	0.53	0.20	0.20	0.19	0.36	0.70
\$25,000	0.51	0.01	0.05	0	0.20	0.54
\$30,000	0.31	0	0.02	0	0.10	0.33
\$35,000	0.25	0	0	0	0.02	0.09
\$40,000	0.22	0	0	0	0	0.02

Note: US subsidy rates take into account the DCTC and CCDF. Baseline specification with logarithmic social welfare function and with $1e = 24$, $\omega = \$5.1$ and $\gamma = 1$. In sensitivity analysis, we recalibrate θ and M by varying the parameter of reference while keeping the other ones at the baseline level. Because z is discrete, we do not always observe a z with earnings level exactly equal to say $5k, 10k, 20k, 25k, 30k, 35k, 40k$. We use linear interpolation to approximate the subsidy rates in between discretized earnings levels where necessary.

Comparing the third and fourth columns of the Table, we note that the subsidy rates in the logarithmic and Rawlsian social welfare functions are very similar and are nearly identical for income levels below \$25,000. The subsidy rates are more variable across the other specifications, with higher optimal subsidy rates for the specifications with high formal child care cost (\$6.40) and with more inelastic labor supply ($\gamma = 2$).

In Table 5, we report the marginal income tax rates computed according to equation (8), which for all cases where $f(y) = \omega$ is optimal, specializes to: $T'(y) = 1 - \frac{1}{z} v'\left(\frac{y}{z}\right) + s'_y(y, \omega) f(y)$. *Despite the steeply decreasing subsidy rates for child care costs, optimal marginal tax rates are positive at all levels of earnings*. The comparison between the third and fourth columns in Table 5 also indicates that, as expected, employed agents tend to face higher marginal income tax rates under the Rawlsian social welfare function compared to the benchmark case. Comparing the rows corresponding to earnings level of \$5,000 in both Tables 4 and 5, we note that low income earners face lower optimal marginal taxes in specifications that display higher optimal child care subsidy rates.

Child Allowances As explained in Section 5.2, the current US tax system does not need to be reformed. We keep actual net taxes faced by a single childless individual, $T^a(y)$ (i.e., federal and SS taxes, EITC, and unemployment benefits), and find the corresponding optimal child

Table 5: **Marginal Income Tax Rates for Employed Mothers**

Earnings	US	Baseline	Rawls	$1e = 34$	$\omega = 6.4$	$\gamma = 2$
\$5,000	-0.32	0.24	0.52	0.31	0.22	0.19
\$10,000	-0.32	0.32	0.47	0.43	0.32	0.40
\$15,000	0.08	0.37	0.41	0.38	0.37	0.40
\$20,000	0.39	0.29	0.41	0.37	0.29	0.34
\$25,000	0.39	0.33	0.42	0.39	0.34	0.39
\$50,000	0.23	0.29	0.33	0.30	0.30	0.29
\$75,000	0.33	0.24	0.26	0.25	0.24	0.13
\$100,000	0.33	0.17	0.18	0.17	0.17	0

Note: US taxes take into account Federal and SS tax rates as well as the EITC rates. Baseline specification with logarithmic social welfare function and with $1e = 24$, $\omega = \$5.1$ and $\gamma = 1$. In sensitivity analysis, we recalibrate θ and M by varying the parameter of reference while keeping the other ones at the baseline level. The table reports the adjusted optimal marginal labor income tax rates for different income levels, computed as the sum of the labor wedges and marginal child care subsidies as in equation (8): $T'(y) = 1 - \frac{1}{z}v'\left(\frac{y}{z} + 1 - \frac{f(y)}{\omega}\right) + s'_y(y, f(y))f(y)$. Because z is discrete, we do not always observe a z with earnings level exactly equal to say $5k, 10k, 15k, 20k, 25k, 50k, 75k, 100k$. We use linear interpolation to approximate the marginal tax rates in between discretized earnings levels where necessary.

allowances, $A(y)$. The net income taxes are computed as in Section 2.

The net income taxes $T^a(y)$, optimal child care subsidy rates $s(y, f(y))$ and child allowances $A(y)$ for the baseline specification are illustrated in Figure 6. Panels (b) and (e) plot the optimal child allowances, $A(y)$, in comparison to the US child allowances (i.e., federal income tax exemptions, EITC and TANF as computed in Section 2). In Panels (c) and (f), we compare the optimal baseline child care subsidy rates to those of the US (i.e., DCTC and CCDF as computed in Table 4).

Qualitatively at least, the US child care subsidy scheme looks very similar to our optimal subsidy scheme with child care subsidy rates declining with earnings. As argued above, the optimal program provides stronger participation incentives compared to the US scheme. The intensive margin incentives provided by the efficient program however are mild, with child care subsidy rates decreasing with income more steeply than those in the US and child care allowances much flatter than those of the US scheme, especially at low and intermediate levels of earnings.

Second Best Implementation with Actual US Tax System

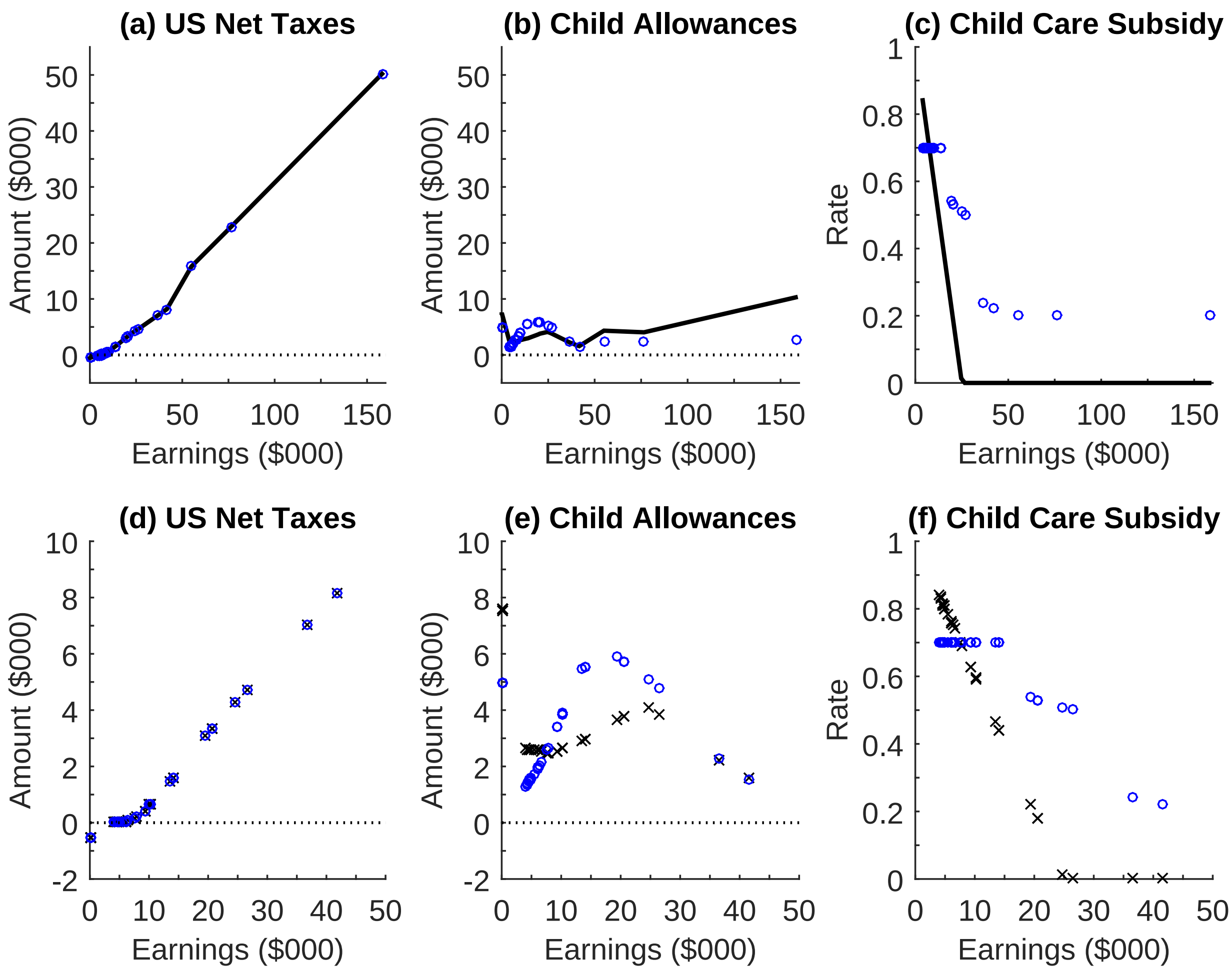

Figure 6: The solid black lines and the black crosses denote the optimal taxes and benefits whereas the blue circles denote the US taxes and benefits. Panel (a) depicts the net income taxes (Federal and SS Taxes, EITC and unemployment benefits) under the actual tax system, $T^a(y)$, faced by a single childless individual. The solid line in Panel (b) depicts the optimal child care subsidy rates, $s(y)$, while the circles denote the combined DCTC and CCDF subsidy rates. The solid line in Panel (c) depicts the corresponding optimal child allowances, $A(y)$, while the circles capture the more generous income tax exemptions, EITC and TANF benefits that a single parent is eligible for in the US relative to a childless individual. The net income taxes, optimal child care subsidy rates and optimal child allowances implement the second best. In the bottom panels (d), (e) and (f), we report the corresponding figures for net taxes, child allowances and child care subsidy rate zoomed for the range of earnings below $50,000.

Further Sensitivity Analysis Since the optimal subsidy rates proposed in Proposition 4 depends on the productivity level z_N, we provide further sensitivity analysis, where we vary the number of wage bins to 25, 100, and 500. We illustrate the optimal child care subsidy rates and child allowances in Figure 8 in Appendix B.5. As can be seen from the Figure, the optimal subsidy rates and allowances do not vary much across wage bin grids. Further details are available in Appendix B.5.

7 Conclusion and Discussion

We provide an efficiency case for child care subsidies in an optimal tax and welfare design problem. We show that optimal child care subsidy rates follow a sliding scale and that the coverage rates should contemplate a kink. These feature are in line with the qualitative features of the existing US scheme. Although child care subsidies incentivize higher work participation, the sliding scale pattern may have disincentive effects on labor supply. To counterbalance such disincentives, marginal labor income taxes are set at lower rates than the labor wedges. Overall, we find that the optimal program provides stronger work participation incentives but milder intensive margin incentives than the US scheme.

A main achievement of this paper is to formulate a flexible model of the design of child care subsidies and to derive a number of properties of the optimal scheme. This might serve to unify a body of literature and to suggest some new results. Despite the complexity of the resulting screening problem, the solution found is remarkably simple and can be explained intuitively. The theory that emerges has a non-local nature.

The model does, however, have a number of limitations. Our assumptions are fairly stringent. First, we have used the assumption of quasilinear preferences. Second, we have conducted the analysis assuming that individuals differ only with respect to their productivities in the primary labor market. Neither seems particularly realistic, although we do not believe that this nullifies the value of our analysis. The considerations that we have uncovered are likely to be important in more general analyses. Third, our assumptions about how the labor market operates are somewhat restrictive. For example, we have abstracted from general equilibrium effects. We do, however, share this limitation with most of the literature on optimal income taxation.

The efficiency case for child care subsidies would still hold for more general non-separable preference specifications although the optimal child care subsidy scheme would depend on the marginal rates of substitution between consumption and effort, possibly loosing the sliding-scale pattern that we find here.[23] Our model is also extendable to heterogeneous quality choices. It is straightforward to see that if quality is verifiable, a tax and subsidy scheme that takes into account quality differences, such as quality related vouchers (Blau, 2003), would be efficient. The problem would be more complicated if quality is not verifiable since formal child care cost would then be a function of both formal child care usage and quality usage. Bastani, Blomquist, and Micheletto (2013) explore the desirability of a refundable tax credit, tax deductability, and public provision of child care in a simple model with two agent's types. Their main focus is precisely on motivating parents to choose higher quality child care. Child quality considerations are an important issue that, we believe, deserves further investigation.

We have also abstracted from dynamic considerations in our model. Following standard arguments, it can be shown that if z does not change over time, the optimal dynamic allocation is a repetition of the static one characterized in this paper (Baron and Besanko, 1984; and Fudenberg and Tirole, 1991, pp. 299-301). With strictly concave preferences in consumption and stochastic z, matters become more complicated and the taxation of savings becomes relevant for redistribution (Abraham, Koehne, and Pavoni, 2015; Ho, 2015a; and Kocherlakota, 2010).

Taking into account the potential human capital accumulation losses or gains from encouraging participation may also be interesting. Blundell et al. (2013), for example, find that single mothers with basic education earn little or no returns to experience while those who are more highly educated have significant returns to experience. The fact that the potential gains from incentivizing participation are unequal across skill groups might have non-trivial implications for the optimal pattern of child care subsidies. We leave these considerations for future research.

[23] With additive separable preferences - $U(c,e) = u(c) - v(e)$, with u concave and v convex - as long as the efficient consumption allocation increases with income, the sliding scale feature will be maintained.

References

Abraham, A, S Koehne, and N Pavoni. 2015. "Optimal Income Taxation when Asset Taxation is Limited." *Journal of Public Economics* forthcomin.

Administration for Children and Families. 2012. "Child Care and Development Fund: Report to Congress FY 2008-2011." *Report* .

Bainbridge, J., M.K. Meyers, and J. Waldfogel. 2003. "Child Care Policy Reform and the Employment of Single Mothers." *Social Science Quarterly* 84 (4):771–791.

Barnett, W. S. 1993. "New Wine in Old Bottles: Increasing the Coherence of Early Childhood Care and Education Policy." *Early Childhood Research Quarterly* 8 (4):519–558.

Baron, D and D Besanko. 1984. "Regulation and Information in a Continuing Relationship." *Information Economics and Policy* 1 (3):447–470.

Bastani, S., S. Blomquist, and L. Micheletto. 2013. "Child Care Subsidies, Quality and the Optimal Income Tax." *Mimeo* .

Beaudry, P., C. Blackorby, and D. Szalay. 2009. "Taxes and Employment Subsidies in Optimal Redistribution Programs." *American Economic Review* 99 (1):216–242.

Besley, T. and S. Coate. 1995. "The Design of Income Maintenance Programmes." *Review of Economic Studies* 62 (2):187–221.

Blau, D. 2003. "Child Care Subsidy Programs." In *Means-Tested Transfer Programs in the United States*, vol. 3, edited by R. Moffitt. University of Chicago Press.

Blau, D. and E. Tekin. 2007. "The Determinants and Consequences of Child Care Subsidies for Single Mothers in the USA." *Journal of Population Economics* 20 (4):719–741.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw. 2013. "Female Labor Supply, Human Capital and Welfare Reform." .

Blundell, R., C. Meghir, and P. Neves. 1993. "Labor Supply and Intertemporal Substitution." *Journal of Econometrics* 59 (1):137–160.

Chetty, R., A. Guren, D. Manoli, and A. Weber. 2011. "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review P&P* 101 (3):471–475.

Child Care Aware of America. 2012. "Parents and the High Cost of Child Care." *Report* .

Choné, P. and G. Laroque. 2011. "Optimal Taxation in the Extensive Model." *Journal of Economic Theory* 146 (2):425–453.

Connelly, R and J Kimmel. 2003. "Marital Status and Full–Time/Part–Tme Work Status in Child Care Choices." *Applied Economics* 35 (7):761–777.

Currie, J. 2001. "Early Childhood Education Programs." *The Journal of Economic Perspectives* 15 (2):213–238.

Domeij, D. and P. Klein. 2013. "Should day care be subsidized?" *The Review of Economic Studies* 80 (2):568–595.

Fudenberg, D and J Tirole. 1991. *Game Theory*. MIT Press.

Gabe, T., B. Lyke, and K. Spar. 2001. "Child Care Subsidies: Federal Grants and Tax Benefits for Working Families." Tech. rep., Congressional Research Service Report.

Golosov, M., M. Troshkin, A. Tsyvinski, and M. Weinzierl. 2013. "Preference Heterogeneity and Optimal Capital Income Taxation." *Journal of Public Economics* 97 (1):160–175.

Guner, N., G. Ventura, and R. Kaygusuz. 2013. "Childcare Subsidies and Household Labor Supply." *Mimeo* .

Han, W. J. and J. Waldfogel. 2001. "The Effect of Child Care Costs on the Employment of Single and Married Mothers." *Social Science Quarterly* 82 (3):552–68.

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

Heckman, J. and T. Macurdy. 1980. "Life Cycle Model of Female Labour Supply." *The Review of Economic Studies* 47 (1):47–74.

Ho, C. 2013. "Welfare Reform and At-Risk Mothers' Labour Supply." *Economic Papers* 32 (2):249–257.

———. 2015a. "Optimal Social Insurance and Universal Day Care." *Mimeo* .

———. 2015b. "Welfare-to-Work Reform and Intergenerational Support: Grandmothers' Response to the 1996 PRWORA." *Journal of Marriage and Family* 77 (2):407–423.

Karoly, L. A., P. W. Greenwood, S. S. Everingham, J. Houbé, M. R. Kilburn, C. P. Rydell, and J. Chiesa. 1998. *Investing in Our Children: What We Know and Don't Know About the Costs and Benefits of Early Childhood Intervenions*. RAND Monograph.

Kimmel, J. 1995. "The Effectiveness of Child-Care Subsidies in Encouraging the Welfare-to-Work Transition of Low-Income Single Mothers." *American Economic Review P&P* 85 (2):271–275.

Kleven, H. J., W. F. Richter, and P. B. Sørensen. 2000. "Optimal Taxation with Household Production." *Oxford Economic Papers* 52 (3):584–594.

Kocherlakota, N. 2010. *The New Dynamic Public Finance*. Princeton University Press.

Kocherlakota, N R. 2004. "Wedges and Taxes." *American Economic Review Papers & Proceedings* 94 (2):109–113.

Laughlin, L. 2010. "Who's Minding the Kids? Child Care Arrangements." Tech. rep.

Mankiw, N. G., M. Weinzierl, and D. Yagan. 2009. "Optimal Taxation in Theory and Practice." *Journal of Economic Perspectives* 23 (4):147–174.

Matthews, S. and J. Moore. 1987. "Monopoly Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening." *Econometrica* 55 (2):441–467.

Minton, S., C. Durham, E. Huber, and L. Giannarelli. 2012. "The CCDF Policies Database Book of Tables: Key Cross-State Variations in CCDF Policies as of October 1, 2011." *OPRE Report* .

National Center for Infant, Toddlers, and Families. 2015. "What's in the Budget for Babies." *Report* .

Rosenbaum, DT and CJ Ruhm. 2007. "Family Expenditures on Child Care." *The BE Journal of Economic Analysis & Policy* 7 (1).

Saez, E. 2002a. "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses." *Quarterly Journal of Economics* 117 (3):1039–1073.

———. 2002b. "The Desirability of Commodity Taxation under Non-Linear Income Taxation and Heterogeneous Tastes." *Journal of Public Economics* 83 (2):217–230.

Tax Policy Center. 2010a. "Earned Income Tax Credit Parameters." *Report* .

———. 2010b. "Quick Facts: Child and Dependent Care Tax Credit." *Report* .

———. 2010c. "State Individual Income Taxes 2010." *Report* .

Tekin, E. 2005. "Child Care Subsidy Receipt, Employment, and Child Care Choices of Single Mothers." *Economics Letters* 89 (1):1–6.

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

———. 2011. "Characteristics and Financial Circumstances of TANF Recipients: FY2010." *Report* .

Werning, I. 2011. "Nonlinear Capital Taxation." *Mimeo* .

Wooldridge, J. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

A Proofs

For ease of exposition, we report the statement in each Lemma and Proposition from the main text (*in italics*) before each proof. We start with a couple of preparatory results.

Claim 1 *In an optimal contract, we have $y(z_1) = 0$; and if $h(z_1) > 0$, then it solves*

$$1 - \frac{1}{\omega} v'(h(z_1)) \geq 0, \tag{10}$$

with equality whenever $v'(1) \geq \omega$. If $v'(1) \leq \omega$, then $h(z_1) = 1$. In addition, if for some z we have $y(z) = 0$, then agent z gets the same allocation as type $z_1 = 0$.

Proof. Since $z_1 = 0$, we must have $y(z_1) = 0$. From the first order conditions of agent $z_1 = 0$ with respect to c and h, we have $v'(h(z_1)) \leq \omega$. Since $v'(0) = 0$ and $\omega > 0$, we have $h(z_1) > 0$. If $v'(1) \leq \omega$, utility can be increased strictly by increasing h whenever $h(z_1) < 1$; Hence, it must be that $h(z_1) = 1$. Consider now type $z > 0$ declaring $\sigma = z_1$. When $y(\sigma) = 0$, all agents have the same preferences over c and h and get the same utility when declaring $\sigma = z_1$. Thus, by DIC, agent z must receive at least the same utility as agent z_1. If $y(z) = 0$, UIC implies the reverse inequality, so that the utility between z and z_1 must be the same. Since agent z_1's problem is strictly concave, the allocation designed for z_1 minimizes the budget cost. Hence, we should use for all z with $y(z) = 0$ the allocation designed for z_1. ■

Claim 2 *Let λ be the multiplier associated to the budget constraint (1). We have $\lambda = \sum_{i=1}^{N} \pi(z_i)\phi(z_i) = \mathbf{E}[\phi]$.*

Proof. This result is shown by a simple variational exercise. Since we can increase $c(z_i)$ by the same amount for all i without violating the incentive constraints, it must be that $\sum_{i=1}^{N} \pi(z_i)\phi(z_i) \leq \lambda$. Since we can also decrease all $c(z_i)$ uniformly in an incentive compatible way, it must be that $\sum_{i=1}^{N} \pi(z_i)\phi(z_i) \geq \lambda$. Combining the two we get the desired equality. ■

A.1 Proof of Lemma 1

Under Assumptions 1, 2 and 3, any solution to the second best problem where only downward incentive constraints are imposed - that is, when the set of conditions (2) is relaxed to be $\sigma \leq z$ - delivers an optimal allocation. In addition, the 'local' downward incentive constraints can be imposed as equalities. Finally, if the upward incentive constraint is binding for two types $z_j < z_k$, then it is optimal for all agents with type $z_i : z_j \leq z_i \leq z_k$ to receive the same allocation (i.e, bunching).

Proof. We want to show that the solution from the relaxed second best problem, where the government maximizes the objective (3) subject to the budget constraint (1) and only the DIC in (2), is the solution to the original problem. In particular, we delete the UIC when the allocation is of employment whereas the unemployed allocation is one of pooling from Claim 1. The problem is a relaxed one, although the set of constraints that we neglect is endogenous to the chosen allocation. The relaxed second best problem is:

$$\max_{c(\cdot),y(\cdot),h(\cdot)} \quad \sum_{i=1}^{N} \pi\left(z_i\right) \phi(z_i) \left[c\left(z_i\right) - v\left(\frac{y\left(z_i\right)}{z_i} + h\left(z_i\right)\right)\right] \qquad \text{(R)}$$

s.t.

$$\sum_{i=1}^{N} \pi\left(z_i\right) c\left(z_i\right) + \omega \leq \sum_{i} \pi\left(z_i\right) \left[y\left(z_i\right) + \omega h(z_i)\right] + M,$$

and for all i with $y(z_i) > 0$:

$$c\left(z_i\right) - v\left(\frac{y\left(z_i\right)}{z_i} + h\left(z_i\right)\right) \geq c\left(z_j\right) - v\left(\frac{y\left(z_j\right)}{z_i} + h\left(z_j\right)\right) \; \forall j < i; \qquad \text{DIC(i)}$$

Finally: $y(z_1) = 0$, and $\forall i$ such that $y(z_i) = 0$, we impose $c(z_i) - v(h(z_i)) = c(z_1) - v(h(z_1))$.

Step 1: We start with a lemma that shows that the double crossing condition described by Matthews and Moore (1987) holds for our framework.

Lemma 2 *Assume that Assumptions 1 and 2 hold. Let $z_- < z_0 < z_+$ be three ordered values of productivity. Let $\bar{w} := (\bar{c}, \bar{y}, \bar{h})$ and $\hat{w} := (\hat{c}, \hat{y}, \hat{h})$ be two allocations, and for all $z > 0$ define:*

$$u(w; z) := c - v\left(\frac{y}{z} + h\right).$$

Suppose that we have

$$u(\bar{w}; z_-) \geq u(\hat{w}; z_-),$$

$$u(\bar{w}; z_+) \geq u(\hat{w}; z_+),$$

but

$$u(\hat{w}; z_0) \geq u(\bar{w}; z_0),$$

with at least one inequality holding as strict. Then **(a)** *$\bar{h} > \hat{h}$, $0 < \bar{y} < \hat{y}$, and $\frac{\bar{y}}{z_*} + \bar{h} > \frac{\hat{y}}{z_*} + \hat{h}$, where $z_* \in (z_-, z_+)$ is the value for which the function $f(\cdot) := u(\hat{w}; \cdot) - u(\bar{w}; \cdot)$ takes the max with respect to z;* **(b)** *if $u(\bar{w}; z_+) > u(\hat{w}; z_+)$, then we have $u(\bar{w}; z_{++}) > u(\hat{w}; z_{++})$ for all $z_{++} > z_+$; and* **(c)** *if $u(\bar{w}; z_-) > u(\hat{w}; z_-)$, then we have $u(\bar{w}; z_{--}) > u(\hat{w}; z_{--})$ for all $z_{--} < z_-$.*

Proof. A graphical representation of the result is reported in Figure 7 for ease of exposition. Let $z_* \in (z_-, z_+)$ be the value for which the function $f(z) := u(\hat{w}; z) - u(\bar{w}; z)$ takes the max. The necessary first

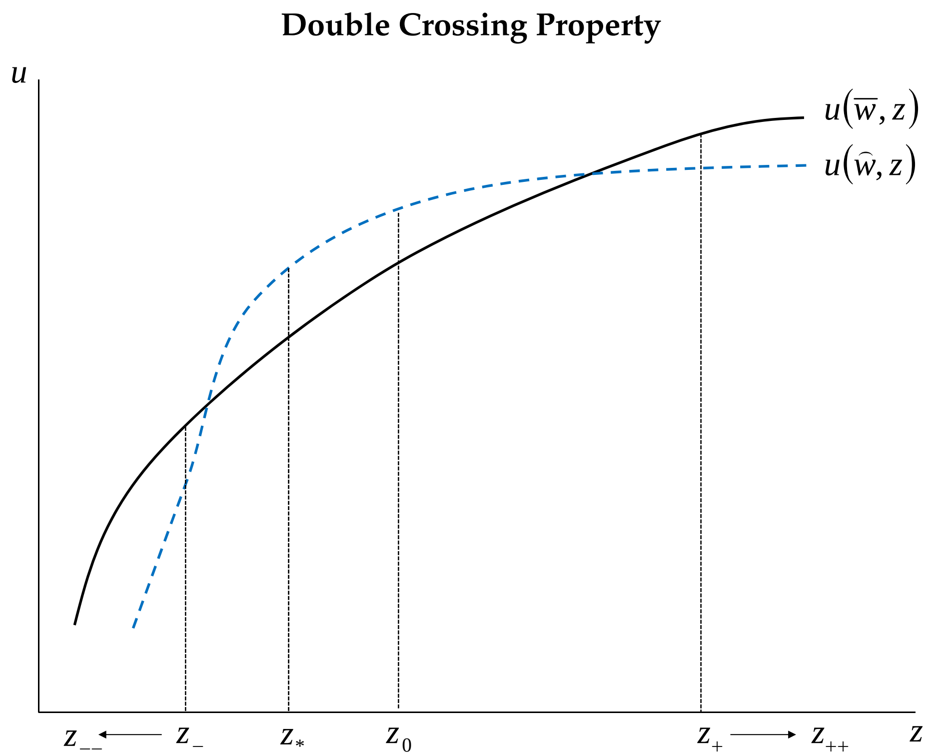

Figure 7: The above figure illustrates the double crossing property of indifference curves.

order condition (FOC) and second order condition (SOC) are respectively:

$$f'(z) = \frac{\hat{y}}{z_*^2} v'\left(\frac{\hat{y}}{z_*} + \hat{h}\right) - \frac{\bar{y}}{z_*^2} v'\left(\frac{\bar{y}}{z_*} + \bar{h}\right) = 0,$$

$$f''(z) \leq 0 \iff \frac{\hat{y}^2}{z_*^4} v''\left(\frac{\hat{y}}{z_*} + \hat{h}\right) - \frac{\bar{y}^2}{z_*^4} v''\left(\frac{\bar{y}}{z_*} + \bar{h}\right) = \frac{\hat{y}}{z_*^4} \frac{v''\left(\frac{\hat{y}}{z_*} + \hat{h}\right)}{v'\left(\frac{\hat{y}}{z_*} + \hat{h}\right)} - \frac{\bar{y}}{z_*^4} \frac{v''\left(\frac{\bar{y}}{z_*} + \bar{h}\right)}{v'\left(\frac{\bar{y}}{z_*} + \bar{h}\right)} \geq 0,$$

where we used the FOC to simplify and rearrange the expression for the SOC. From the FOC, if either $\bar{y} = 0$ or $\hat{y} = 0$, then it must be that both $\bar{y}$ and $\hat{y}$ are equal to 0. But then, from Claim 1, this would imply that $\hat{h} = \bar{h}$, so that the assumptions of the lemma would not be satisfied. So we can safely presume that both $\bar{y}$ and $\hat{y}$ are positive so that the expression for the SOC is well defined.

(a) Suppose that $\bar{h} < \hat{h}$ and recall that v is strictly convex. Then from the FOC, it must be that $\bar{y} > \hat{y}$. Since by Assumption 1, the ratio $\frac{v''}{v'}$ decreases in the argument, the SOC can be satisfied only if $\frac{\bar{y}}{z_*} + \bar{h} > \frac{\hat{y}}{z_*} + \hat{h}$. But then, the FOC would not be satisfied. So it must be that $\bar{h} \geq \hat{h}$ and as a consequence of the FOC, $\bar{y} \leq \hat{y}$ and $\frac{\bar{y}}{z_*} + \bar{h} \geq \frac{\hat{y}}{z_*} + \hat{h}$. We can exclude the equality since we assumed that at least one inequality is strict, while if $\bar{h} = \hat{h}$ we must have $\bar{y} = \hat{y}$ and all types will be indifferent between the two allocations. We hence obtain the inequalities as stated in (a).

(b) Assume that $u(\bar{w}; z_+) > u(\hat{w}; z_+)$ and that for a $z_{++} > z_+ > z_*$ we have instead $u(\bar{w}; z_{++}) \leq u(\hat{w}; z_{++})$. Then, there must be a minimum of f in the interval $(z_*, z_{++}]$. It is easy to see, by reverting the SOC inequality in the previous set of necessary conditions, that FOC and SOC would imply that $\hat{h} \geq \bar{h}$ and $\hat{y} \leq \bar{y}$, thereby contradicting the result in (a). Result **(c)** is shown symmetrically. ∎

Step 2: We can now start the core proof of Lemma 1. We use an induction argument. Denote the set of upward incentive constraints associated with mimicking agent i:

$$c\left(z_j\right) - v\left(\frac{y\left(z_j\right)}{z_j} + h\left(z_j\right)\right) \geq c\left(z_i\right) - v\left(\frac{y\left(z_i\right)}{z_j} + h\left(z_i\right)\right) \; \forall j < i. \qquad \text{UIC(i)}$$

First, note that UIC(1) is empty. Note also that since z_1=0, if $y(z_2) > 0$, then the UIC associated with z_1 mimicking z_2 would be satisfied as type $z_1 = 0$ agent would have an infinite cost of effort (since v is convex, we can bound the derivative downwards: $\lim_{e \to \infty} v(e) = \infty$). If $y(z_2) = 0$, then Claim 1 implies that z_2 must receive the same allocation as that of agent z_1 so that both DIC and UIC between z_1 and z_2 are satisfied. In addition, if $y(z_2) > 0$, then the DIC between z_2 and z_1 must also be binding. Otherwise, it would be possible to find a small enough $\epsilon > 0$ such that decreasing $c(z_i)$ for all $i > 1$ by ϵ and increasing $c(z_1)$ by $\frac{\sum_{i=2}^{N} \pi(z_i)}{\pi(z_1)} \epsilon$, would leave the budget constraint unchanged. This consumption perturbation, however, would be incentive compatible as long as the DIC between z_2 and z_1 is slack and ϵ is small enough. Note indeed that each agent $z_i > z_2$ receives the same utility by mimicking z_1. Recall that we impose DIC for all agents. A slack DIC between z_2 and z_1 implies that for all $z_i > z_2$ the DIC of agent i mimicking agent z_1 would also be slack. Finally, since consumption for all $z_i > z_1$ change by the same amount, incentive constraints are not affected among agents $i > 1$. This perturbation would weakly increases welfare since $\phi(z_1) \geq \mathbf{E}[\phi]$, by Assumption 3, hence generating a contradiction to optimality. In summary, we have shown that for $i = 1, 2$ the LDIC is binding and all UIC(i) constraints are satisfied for all $i \leq 2$. This is our starting point for the induction argument.

Now let $1 < k < N$ and assume that all UIC(i) constraints are satisfied for all $i \leq k$. We will show that in the relaxed problem (R) the LDIC must be satisfied with equality. Fix $1 < k < N$ and suppose that in the optimal contract we have:

$$c\left(z_{k+1}\right) - v\left(\frac{y\left(z_{k+1}\right)}{z_{k+1}} + h\left(z_{k+1}\right)\right) > c\left(z_k\right) - v\left(\frac{y\left(z_k\right)}{z_{k+1}} + h\left(z_k\right)\right).$$

We now show that, *if the local LDIC is slack and the induction hypothesis is true, then none of the non-local downward constraints can be binding*. Suppose that we have $z_j < z_k < z_{k+1}$ such that the LDIC between z_k and z_{k+1} is slack while the DIC between z_j and z_{k+1} is binding. Then, it must be that:

$$c\left(z_{k+1}\right) - v\left(\frac{y\left(z_{k+1}\right)}{z_{k+1}} + h\left(z_{k+1}\right)\right) = c\left(z_j\right) - v\left(\frac{y\left(z_j\right)}{z_{k+1}} + h\left(z_j\right)\right) > c\left(z_k\right) - v\left(\frac{y\left(z_k\right)}{z_{k+1}} + h\left(z_k\right)\right). \quad (11)$$

On the other hand, from the DIC we have:

$$c\left(z_k\right)-v\left(\frac{y\left(z_k\right)}{z_k}+h\left(z_k\right)\right)\geq c\left(z_j\right)-v\left(\frac{y\left(z_j\right)}{z_k}+h\left(z_j\right)\right),$$

while from the UIC (which, by the inductive hypothesis, are assumed to be satisfied for $j\leq k$) we have:

$$c\left(z_j\right)-v\left(\frac{y\left(z_j\right)}{z_j}+h\left(z_j\right)\right)\geq c\left(z_k\right)-v\left(\frac{y\left(z_k\right)}{z_j}+h\left(z_k\right)\right).$$

We thus have the conditions to apply Lemma 2, where the three ranked types are $z_j<z_k<z_{k+1}$ and the supposedly optimal bundle for type z_k takes the role of bundle $(\hat{c},\hat{y},\hat{h})$ while the bundle for type z_j plays the role of the $(\bar{c},\bar{y},\bar{h})$ bundle in the Lemma. Lemma 2(b) implies that for all $z_{++}>z_{k+1}$,

$$c\left(z_j\right)-v\left(\frac{y\left(z_j\right)}{z_{++}}+h\left(z_j\right)\right)>c\left(z_k\right)-v\left(\frac{y\left(z_k\right)}{z_{++}}+h\left(z_k\right)\right).$$

But then from DIC, we have that all $z_{++}>z_{k+1}$ prefer their own bundle to that of agent z_j. Hence, we have:

$$c\left(z_{++}\right)-v\left(\frac{y\left(z_{++}\right)}{z_{++}}+h\left(z_{++}\right)\right)\geq c\left(z_j\right)-v\left(\frac{y\left(z_j\right)}{z_{++}}+h\left(z_j\right)\right)>c\left(z_k\right)-v\left(\frac{y\left(z_k\right)}{z_{++}}+h\left(z_k\right)\right).$$

So, no DIC is binding in terms of mimicking z_k. The first order conditions in the relaxed problem will therefore be those of full information. In particular,

$$z_k=v'\left(\frac{y\left(z_k\right)}{z_k}+h\left(z_k\right)\right).$$

At the same time, since z_{k+1} has a binding constraint with z_j, we have:

$$z_j\geq v'\left(\frac{y\left(z_j\right)}{z_j}+h\left(z_j\right)\right).$$

Since $z_k>z_j$, these conditions imply that:

$$v'\left(\frac{y\left(z_k\right)}{z_k}+h\left(z_k\right)\right)>v'\left(\frac{y\left(z_j\right)}{z_j}+h\left(z_j\right)\right). \tag{12}$$

Now, consider the DIC between z_k and z_j:

$$c(z_k) - v\left(\frac{y(z_k)}{z_k} + h(z_k)\right) \geq c(z_j) - v\left(\frac{y(z_j)}{z_k} + h(z_j)\right).$$

Since v is convex and $z_k > z_j$, we also have $v'\left(\frac{y(z_k)}{z_k} + h\left(z_k\right)\right) > v'\left(\frac{y(z_j)}{z_k} + h\left(z_j\right)\right)$ from inequality (12), which together with the DIC implies that $c(z_k) > c(z_j)$.

From (11), since $c\left(z_j\right) - v\left(\frac{y(z_j)}{z_{k+1}} + h\left(z_j\right)\right) > c\left(z_k\right) - v\left(\frac{y(z_k)}{z_{k+1}} + h\left(z_k\right)\right)$, it must therefore be that:

$$v\left(\frac{y(z_k)}{z_{k+1}} + h(z_k)\right) > v\left(\frac{y(z_j)}{z_{k+1}} + h(z_j)\right) \iff \frac{y(z_k)}{z_{k+1}} + h(z_k) > \frac{y(z_j)}{z_{k+1}} + h(z_j). \tag{13}$$

Since Lemma 2(a) implies that $y(z_k) > y(z_j)$ and $h(z_k) < h(z_j)$, inequality (13) implies that $\frac{y(z_k)}{z} + h(z_k) > \frac{y(z_j)}{z} + h(z_j)$ for all $z \leq z_{k+1}$. On the other hand, Lemma 2(a) also implies that $\frac{y(z_k)}{z_*} + h(z_k) < \frac{y(z_j)}{z_*} + h(z_j)$ for some $z_j < z_* < z_{k+1}$. This is hence a contradiction. It must therefore be that if LDIC for z_{k+1} is slack, then all DIC for z_{k+1} must also be slack.

But then, if the DIC for agent z_{k+1} mimicking a lower type are slack, we can find an incentive compatible $\epsilon > 0$, such that we can increase $c(z_i)$ for all $i \neq k+1$ by ϵ and decrease $c(z_{k+1})$ by $\frac{\sum_{i \neq k+1} \pi(z_i)}{\pi(z_{k+1})}\epsilon$, such that the budget constraint remains the same. This will be incentive compatible since the LDIC between z_{k+1} and z_k is slack by assumption and consumption for all $i \neq k+1$ increase by the same amount. Welfare changes by the amount

$$\left[\sum_{i \neq k+1} \pi(z_i)\phi(z_i) - \sum_{i \neq k+1} \pi(z_i)\phi(z_{k+1})\right]\epsilon \geq 0,$$

where the inequality is implied by Assumption 3. Thus, it must be that the LDIC are binding.

We now show that binding LDIC implies that the UIC are satisfied. Note that the binding LDIC for z_{k+1} mimicking z_k implies:

$$y(z_{k+1}) + \omega h(z_{k+1}) - c(z_{k+1}) \geq y(z_k) + \omega h(z_k) - c(z_k).$$

This must be true, otherwise, the budget constraint could be relaxed (strictly) by replacing allocation $(y(z_{k+1}), h(z_{k+1}), c(z_{k+1}))$ with allocation $(y(z_k), h(z_k), c(z_k))$. Namely, we can give to agent z_{k+1} the contract designed for agent z_k. By eliminating one contract, all incentive constraints will remain satisfied and agent z_{k+1}'s utility will be the same as the agent is indifferent between the two allocations. By the induction assumption, since all LDIC are satisfied with equality, we have that $y(z_i) + \omega h(z_i) - c(z_i)$ *weakly increases with* z_i for $i = 1, \ldots, k+1$.

Now, assume that some of the UIC(k+1) are not satisfied. Namely, for some $1 < j \leq k$ (recall that the UIC(1) is empty), we have:

$$c\left(z_{k+1}\right) - v\left(\frac{y\left(z_{k+1}\right)}{z_j} + h\left(z_{k+1}\right)\right) > c\left(z_j\right) - v\left(\frac{y\left(z_j\right)}{z_j} + h\left(z_j\right)\right).$$

Then, it must be that for such z_j we have:

$$y(z_{k+1}) + \omega h(z_{k+1}) - c(z_{k+1}) < y(z_j) + \omega h(z_j) - c(z_j).$$

Otherwise, we could replace allocation $(y(z_j), h(z_j), c(z_j))$ with $(y(z_{k+1}), h(z_{k+1}), c(z_{k+1}))$. The budget constraint would be weakly relaxed, the incentive constraints remain satisfied, and agent z_j would have higher utility. Hence welfare would increase (strictly). But this provides a contradiction to the monotonicity obtained from the binding LDIC. Hence, it must be that all UIC(k+1) are satisfied.

Finally, *we show that bunching may occur and characterize when this happens.* Claim 1 shows that if $y(z_i) = y(z_j) = 0$, then UIC are trivially satisfied and bunching arises. So, assume that $z_i > z_j$ and $y(z_i) > 0$. If the UIC between z_j vs z_i is binding, it must be that $y(z_i) + \omega h(z_i) - c(z_i) \leq y(z_j) + \omega h(z_j) - c(z_j)$. Otherwise, we can eliminate the allocation for agent z_j and give agent z_j the allocation now in place for agent z_i. This would keep welfare the same as UIC is binding and also relax the budget constraint. But since the DIC are binding for all $k \leq i$ and $i > j$, the argument we made above implies that the reverse inequality must also be true. Thus, it much be that $y(z_i) + \omega h(z_i) - c(z_i) \leq y(z_j) + \omega h(z_j) - c(z_j)$. Let's now look at type z_{i-1}. The previous argument implies that $y(z_i) + \omega h(z_i) - c(z_i) = y(z_{i-1}) + \omega h(z_{i-1}) - c(z_{i-1}) = y(z_j) + \omega h(z_j) - c(z_j)$. Recall that the binding LDIC between z_i and z_{i-1} implies that agent z_i is indifferent between the allocation designed for him and the allocation designed for z_{i-1}. So, we can eliminate the allocation designed for him and use the allocation designed for z_{i-1} instead. As usual, this will keep the budget and welfare the same, and possibly relax the incentive constraints. This same argument can be done till agent z_j. Hence a bunching allocation among these agents would be optimal when UIC is binding. ■

A.2 Proof of Proposition 1

Under Assumptions 1, 2 and 3, we have:

- **(a)** *The 'net surplus' $y^*(z) + \omega h^*(z) - c^*(z)$ is non-decreasing in z;*
- **(b)** *Utility of agents in equilibrium $V^*(z|z)$ is non-decreasing in z, and strictly increasing between any two levels $z_{i+1} > z_i$ when $y^*(z_i) > 0$.*
- **(c)** *For all z, $h^*(z) \leq 1$.*

Proof. We omit the superscript * on the optimal allocation for notational simplicity.
(a) The monotonicity property of the net surplus has been shown in the proof of Lemma 1, using the fact that LDIC are satisfied with equality.
(b) We know from Claim 1 that if $y(z_i) = y(z_{i-1}) = 0$, then we have pooling so that z_i and z_{i-1} will get the same utility. Now, suppose that $y(z_i) > y(z_{i-1}) = 0$. Recall that $z_1 = 0$ and $y(z_1) = 0$. Since LDIC implies

$$c\left(z_1\right) - v\left(h\left(z_1\right)\right) \leq c\left(z_i\right) - v\left(\frac{y\left(z_i\right)}{z_i} + h\left(z_i\right)\right),$$

utility must be weakly increasing. Now, assume that the lower type has $y(z_{i-1}) > 0$. Then DIC implies that

$$c\left(z_i\right) - v\left(\frac{y\left(z_i\right)}{z_i} + h\left(z_i\right)\right) \geq c\left(z_{i-1}\right) - v\left(\frac{y\left(z_{i-1}\right)}{z_i} + h\left(z_{i-1}\right)\right) > c\left(z_{i-1}\right) - v\left(\frac{y\left(z_{i-1}\right)}{z_{i-1}} + h\left(z_{i-1}\right)\right),$$

where the first inequality uses the incentive compatibility constraint while the second inequality uses the fact that $z_i > z_{i-1}$ and $y(z_{i-1}) > 0$.
(c) Suppose that we have $h(z_i) > 1$ for some i. Since the marginal return to providing household child care beyond child care needs is zero while the marginal cost is positive, we can reduce both $h(z_i)$ and $c(z_i)$ so that the utility of agent z_i is unchanged (if we denote $\hat{c}$ and $\hat{h}$ as the new values, we have: $c(z_i) - \hat{c} = v\left(\frac{y(z_i)}{z_i} + h(z_i)\right) - v\left(\frac{y(z_i)}{z_i} + \hat{h}\right)$), and relax the budget constraint (which can subsequently be translated to an increase in welfare by a uniform increase in consumption). The incentives of agent z_i to mimic lower type agents are unchanged since utility of agent z_i is unchanged. We also need to show that such a change weakly relaxes the DIC of higher types. When $y(z_i) = 0$, incentives are unchanged since higher type agents get the same utility as agent z_i when mimicking type z_i. When $y(z_i) > 0$, the convexity of v implies that higher types will now get a lower utility when pretending to be type z_i. This is so since for $z > z_i$, we have $v\left(\frac{y(z_i)}{z} + h(z_i)\right) - v\left(\frac{y(z_i)}{z} + \hat{h}\right) < v\left(\frac{y(z_i)}{z_i} + h(z_i)\right) - v\left(\frac{y(z_i)}{z_i} + \hat{h}\right)$ by the convexity of v. ∎

A.3 Proof of Proposition 2

Under Assumptions 1, 2, and 3, we have:

(a) *Unemployment: Recall that $z_1 = 0$. We have $y^*(z_1) = 0$ and $h^*(z_1) > 0$, where*

$$1 - \frac{1}{\omega} v'\left(h^*(z_1)\right) \geq 0, \tag{14}$$

with equality whenever $v'(1) \geq \omega$. If $v'(1) \leq \omega$, then $h^*(z_1) = 1$. In addition, for all z such that $y^*(z) = 0$, type z gets the same allocation as type z_1.

(b) *Low productivity: Let $z \leq \omega$. We have $h^*(z) > 0$, and if $y^*(z) > 0$, then $h^*(z) = 1$.*

(c) *Segmentation: If $y^*(z) > 0$, then $y^*(z') > 0$ for all $z' > z$.*

(d) *Monotonicity: Let $z' > z$ for which we have no bunching. If $h^*(z') \leq h^*(z)$, then $y^*(z') > y^*(z)$; and if $y^*(z') \leq y^*(z)$, then $h^*(z') > h^*(z)$.*

(e) *Wedges for the employed: Let z_i be such that $y^*(z_i) > 0$. Then labor wedges are non-negative:*

$$1 - \frac{1}{z_i} v'(e^*(z_i)) \geq 0; \tag{15}$$

If, in addition, $h^*(z_i) > 0$, then the child care wedges are also non-negative:

$$1 - \frac{1}{\omega} v'(e^*(z_i)) \geq 0. \tag{16}$$

Both wedges are strictly positive whenever $\phi(z_{i+1}) < \mathbf{E}[\phi]$.

For $i = N$, the labor wedge is zero and $h^*(z_N) = 0$.

Proof. We omit the superscript * on the optimal allocation for notational simplicity.

(a) This result has been shown in Claim 1.

(b) If $y(z) = 0$, then agent z gets the unemployed allocation and provides $h(z) = h(0) > 0$, as shown in Claim 1. Now, suppose that for $z \leq \omega$, $y(z) > 0$ but $h(z) < 1$. We show that we can reduce y and increase h so as to keep agent z's utility constant without violating any DIC nor the budget constraint. The budget constraint can only improve since keeping agent's $z \leq \omega$ utility constant implies a change $\Delta y + z\Delta h = 0 \leq \Delta y + \omega \Delta h$, where the last quantity is the change in the budget constraint. The DIC are not affected since all $z' > z$ will now find the new allocation less attractive than before (note that $\Delta h(z) = \frac{\Delta y(z)}{z} \geq \frac{\Delta y(z)}{z'}$ so that type z' mimicking z will now face higher effort cost). This change would strictly improve welfare since we can uniformly redistribute the increase in the budget, $\Delta y + \omega \Delta h$, among all types without altering incentives.

(c) We want to show that if for z_i we have $y(z_i) > 0$, then it must be that $y(z_n) > 0$ for all $n > i$ (we know that $y(z_1) = 0$, so $i > 1$). Suppose that for $k > i > j$ we have both $y(z_j) = y(z_k) = 0$ and $y(z_i) > 0$. From Lemma 1(ii), we know that utility between $i+1$ and i must be strictly increasing since $y(z_i) > 0$. We also know that utility is weakly increasing in type from the DIC. So, the utility of agent $z_k \geq z_{i+1}$ must be strictly larger than the utility of agent $z_j \leq z_i$. But we know that we have pooling among the unemployed, and they all receive the same utility, so that z_k and z_j must have the same utility if $y(z_j) = y(z_k) = 0$. Hence, we get a contradiction.

(d) Since the case where $y(z_{i-1}) = 0$ is trivial to show, we show it for the case where $y(z_{i-1}) > 0$ (hence $i > 1$). From (c) above, we have $y(z_i) > 0$ as well. Using the DIC and UIC between z_i and z_{i-1} (recall that the local DIC binds):

$$\begin{aligned} c(z_i) - v\left(\frac{y(z_i)}{z_i} + h(z_i)\right) &= c(z_{i-1}) - v\left(\frac{y(z_{i-1})}{z_i} + h(z_{i-1})\right) \\ c(z_{i-1}) - v\left(\frac{y(z_{i-1})}{z_{i-1}} + h(z_{i-1})\right) &\geq c(z_i) - v\left(\frac{y(z_i)}{z_{i-1}} + h(z_i)\right). \end{aligned}$$

Adding the two inequalities together and rearranging, we get:

$$v\left(\frac{y\left(z_i\right)}{z_{i-1}}+h\left(z_i\right)\right)-v\left(\frac{y\left(z_i\right)}{z_i}+h\left(z_i\right)\right)\geq v\left(\frac{y\left(z_{i-1}\right)}{z_{i-1}}+h\left(z_{i-1}\right)\right)-v\left(\frac{y\left(z_{i-1}\right)}{z_i}+h\left(z_{i-1}\right)\right).$$

By the first fundamental theorem of calculus, this implies that:

$$\int_{z_{i-1}}^{z_i} v'\left(\frac{y\left(z_i\right)}{s}+h\left(z_i\right)\right)\frac{y\left(z_i\right)}{s^2}ds\leq\int_{z_{i-1}}^{z_i} v'\left(\frac{y\left(z_{i-1}\right)}{s}+h\left(z_{i-1}\right)\right)\frac{y\left(z_{i-1}\right)}{s^2}ds.$$

If $h\left(z_i\right)>h\left(z_{i-1}\right)$, then convexity of $v\left(.\right)$ and $y\left(z_i\right)\geq y\left(z_{i-1}\right)>0$ would imply that the integrand on the left hand side is everywhere larger than the right hand side, a contradiction. Thus, it must be that $y\left(z_i\right)<y\left(z_{i-1}\right)$. The equivalent result says that if $y\left(z_i\right)\geq y\left(z_{i-1}\right)$ then $h\left(z_i\right)\leq h\left(z_{i-1}\right)$. Since the index i was generic, we have shown monotonicity. Clearly, if there is no bunching we can show the stronger result: If $h\left(z_i\right)\geq h\left(z_{i-1}\right)$, then $y\left(z_i\right)<y\left(z_{i-1}\right)$ and the equivalent statement: If $y\left(z_i\right)\geq y\left(z_{i-1}\right)$, then $h\left(z_i\right)<h\left(z_{i-1}\right)$.

(e) Recall the relaxed problem (R). From Lemma 1, we can charaterize the problem by focusing on problem (R). For z_i such that $y(z_i)>0$, the first order conditions with respect to consumption, earnings and household child care, are given by:

$$\begin{aligned}
c\left(z_i\right) &: \left[\pi\left(z_i\right)\phi(z_i)+\delta\left(z_i\right)\right]v'\left(\tfrac{y(z_i)}{z_i}+h(z_i)\right) &=& \left[\textstyle\sum_{k>i}\mu_i(z_k)+\lambda\pi\left(z_i\right)\right]v'\left(\tfrac{y(z_i)}{z_i}+h(z_i)\right),\\
y\left(z_i\right) &: \left[\pi\left(z_i\right)\phi(z_i)+\delta(z_i)\right]v'\left(\tfrac{y(z_i)}{z_i}+h(z_i)\right) &=& \textstyle\sum_{k>i}\mu_i(z_k)\tfrac{z_i}{z_k}v'\left(\tfrac{y(z_i)}{z_k}+h(z_i)\right)+\lambda\pi\left(z_i\right)z_i,\\
h\left(z_i\right) &: \left[\pi\left(z_i\right)\phi(z_i)+\delta(z_i)\right]v'\left(\tfrac{y(z_i)}{z_i}+h(z_i)\right) &\geq& \textstyle\sum_{k>i}\mu_i(z_k)v'\left(\tfrac{y(z_i)}{z_k}+h(z_i)\right)+\lambda\pi\left(z_i\right)\omega,
\end{aligned}$$

where $\mu_j(z_i)\geq 0$ is the Kuhn-Tucker multiplier associated with the DIC guaranteeing that agent z_i does not mimic type $z_j<z_i$. For all i, we defined $\delta\left(z_i\right):=\sum_{j<i}\mu_j(z_i)$ and multiplied the first condition by $v'\left(\frac{y(z_i)}{z_i}+h(z_i)\right)$ and the second one by $z_i>0$. Clearly, the first order conditions for y and h are satisfied with equality when we have an interior solution for them.

Substituting the first order condition with respect to $c(z_i)$ into those with respect to $y(z_i)$ and $h(z_i)$, and rearranging, we get:

$$\lambda\pi(z_i)z_i=\lambda\pi(z_i)v'\left(\frac{y(z_i)}{z_i}+h(z_i)\right)+\sum_{k>i}\mu_i(z_k)\left[v'\left(\frac{y(z_i)}{z_i}+h(z_i)\right)-\frac{z_i}{z_k}v'\left(\frac{y(z_i)}{z_k}+h(z_i)\right)\right],$$

and

$$\lambda\pi(z_i)w\leq\lambda\pi(z_i)v'\left(\frac{y(z_i)}{z_i}+h(z_i)\right)+\sum_{k>i}\mu_i(z_k)\left[v'\left(\frac{y(z_i)}{z_i}+h(z_i)\right)-v'\left(\frac{y(z_i)}{z_k}+h(z_i)\right)\right],$$

with equality if $h(z_i)>0$. The term in square brackets are strictly positive since $z_k>z_i$ and v is convex.

It is now easy to see how we can get the wedges.

We now want to show that when $\phi(z_{i+1}) < \lambda = \mathbf{E}[\phi]$ (where the last equality is from Claim 2), then some DIC must be binding with $\mu_i(z_k) > 0$ for some $k > i$.

Suppose that all $\mu_i(z_k)$ are nil. Then, from the first order conditions with respect to $y(z_i)$ and $h(z_i)$, we have $\lambda\pi(z_i) z_i \geq \lambda\pi(z_i)\omega$, which implies $z_i \geq \omega$. Excluding the knife edge case of $z_i = \omega$,[24] the first order conditions imply $h(z_i) = 0$ and:

$$z_i = v'\left(\frac{y(z_i)}{z_i}\right).$$

In addition, for all $z_j < z_i$ for which $y(z_j) > 0$, we have:

$$z_j \geq v'\left(\frac{y(z_j)}{z_j} + h(z_j)\right);$$

and if $y(z_j) = 0$, we have $h(z_j) > 0$ and:

$$\omega \geq v'\left(h(z_j)\right).$$

Since both $z_i > z_j$ and $z_i \geq \omega$, the two inequalities imply:

$$v'\left(\frac{y(z_i)}{z_i}\right) > v'\left(\frac{y(z_j)}{z_j} + h(z_j)\right) \quad \text{or} \quad v'\left(\frac{y(z_i)}{z_i}\right) \geq v'\left(h(z_j)\right).$$

Convexity of v and $h(z_j) \geq 0$ imply $y(z_j) < y(z_i)$ as a consequence of the first inequality, and $y(z_j) = 0 < y(z_i)$ as a consequence of the second. Allowing for zeros, we can summarize the above conditions by:

$$\frac{y(z_i)}{z_i} > \frac{y(z_j)}{z_j} + h(z_j) \quad \text{and} \quad y(z_j) < y(z_i),\ h(z_j) \geq h(z_i) = 0.$$

In addition, since $z_{i+1} > z_i > z_j$ we have $\frac{y(z_i)}{z_i} > \frac{y(z_j)}{z_{i+1}} + h(z_j)$. From Proposition 1(ii) since utility is increasing, we have:

$$c_i - v\left(\frac{y(z_i)}{z_i}\right) > c_j - v\left(\frac{y(z_j)}{z_j} + h(z_j)\right).$$

This, together with $\frac{y(z_i)}{z_i} > \frac{y(z_j)}{z_{i+1}} + h(z_j)$ implies $c(z_i) > c(z_j)$.

Now, if we look at the first order condition for $c(z_{i+1})$, since by assumption, $\phi(z_{i+1}) < \lambda$, we must have $\delta(z_{i+1}) > 0$. As a consequence, by the definition of $\delta(z_{i+1})$, some DIC is binding for agent $i+1$,

[24] If $z_i = \omega$, point (b) above implies that in this case, $y(z_i) > 0$ implies $h(z_i) = 1$. It would also imply either $y(z_j) = 0$ or $h(z_j) = 1$ for all $z_j < z_i$. Hence $y(z_i) > y(z_j)$ and the proof would follow the same line as we do here assuming that $z_i > \omega$.

with positive multiplier: $\mu_j(z_{i+1}) > 0$ for some $j < i$ (by assumption, $\mu_i(z_{i+1}) = 0$). Consider such a j. Since the LDIC for z_{i+1} is binding (and by assumption for such j the non-local DIC z_{i+1} vs z_j binds), we have:

$$\begin{aligned} c_{i+1} - v\left(\frac{y(z_{i+1})}{z_{i+1}}\right) &= c_i - v\left(\frac{y(z_i)}{z_{i+1}}\right), \\ &= c_j - v\left(\frac{y(z_j)}{z_{i+1}} + h(z_j)\right). \end{aligned}$$

Since we saw above that $c(z_i) > c(z_j)$, it must be that:

$$v\left(\frac{y(z_i)}{z_{i+1}}\right) > v\left(\frac{y(z_j)}{z_{i+1}} + h(z_j)\right) \iff \frac{y(z_i)}{z_{i+1}} > \frac{y(z_j)}{z_{i+1}} + h(z_j). \tag{17}$$

Now, define:

$$\Delta(z) := c(z_i) - v\left(\frac{y(z_i)}{z}\right) - \left[c(z_j) - v\left(\frac{y(z_j)}{z} + h(z_j)\right)\right].$$

Using again the fact that both the LDIC for z_{i+1} and the non-local DIC z_{i+1} vs z_j bind, we have that $\Delta(z_{i+1}) = 0$. Moreover, DIC (recall $j < i$) implies $\Delta(z_i) \geq 0$. At the same time, deriving with respect to z, we have:

$$\Delta'(z) = \frac{y(z_i)}{z^2} v'\left(\frac{y(z_i)}{z}\right) - \frac{y(z_j)}{z^2} v'\left(\frac{y(z_j)}{z} + h(z_j)\right).$$

Since $y(z_i) > y(z_j)$, *if we show that for all $z \in [z_i, z_{i+1}]$ we have $v'\left(\frac{y(z_i)}{z}\right) > v'\left(\frac{y(z_j)}{z} + h(z_j)\right)$, then we would be done*. This is so since the above inequality implies that $\Delta'(z) > 0$ for $z \in [z_i, z_{i+1}]$, contradicting the fact that $\Delta(z_i) \geq \Delta(z_{i+1}) = 0$. This would hence mean that the initial assumption was false, namely we must have some $\mu_k(z_i) > 0$. Recall that from (17) we have $\frac{y(z_i)}{z_{i+1}} > \frac{y(z_j)}{z_{i+1}} + h(z_j)$. Since $y(z_i) > y(z_j)$, then $\frac{y(z_i)}{z} > \frac{y(z_j)}{z} + h(z_j)\ \forall z \leq z_{i+1}$ as desired.

We have hence shown that for z_i with $y(z_i) > 0$ and $\phi(z_{i+1}) < \lambda = \mathbf{E}[\phi]$ (thus, $z_i < z_N$):

$$z_i > v'(e(z_i)).$$

Now, suppose that for z_i we have both $y(z_i) > 0$ and $h(z) > 0$. If $z_i > \omega$, then the first order conditions exclude the possibility that all multipliers $\mu_i(z_k)$ are zero. On the other hand, we saw above that if all multipliers $\mu_i(z_k)$ are zero and $y(z_i) > 0$, then it must be that $z_i \geq \omega$. If we exclude the case $z_i = \omega$, we hence have that:[25]

$$\omega > v'(e(z_i)).$$

[25]The case were all multipliers are zero and $z_i = \omega$ also has $h(z_i) = 1$ from point (b) above. If $v'(1) > \omega$, then it cannot be. This case hence can only happen when $v'(1) \leq \omega$. For this case, we can follow the same line of proof as before to derive the wedge for $y(z_i)$ to obtain a contradiction (recall that if $z_j < z_i$, then $z_j < \omega$).

The standard no distortion at the top result is easily obtained from the first order conditions as no DIC exists for this agent. Since $z_N > \omega$, we must have $h(z_N) = 0$ and:

$$1 - \frac{1}{z_N} v' \left(\frac{y(z_N)}{z_N} \right) = 0 \Rightarrow 1 - \frac{1}{\omega} v' \left(\frac{y(z_N)}{z_N} \right) < 0.$$

■

A.4 Proof of Proposition 3

Let $f^*(\sigma)$ be the optimal formal child care cost associated with the constrained optimal $h^*(\sigma)$. The following subsidy rates and transfers implement the constrained optimum.
(a) For employed agents, we have:

$$\text{If} \quad \sigma \notin Z_0^*, \quad \text{then} \quad s(\sigma, f) = \begin{cases} \left(1 - \frac{1}{\omega} v' \left(\frac{y^*(\sigma)}{z_N} + h^*(\sigma) \right) \right)^+ & \text{if} \quad f \leq f^*(\sigma); \\ \left(1 - \frac{1}{\omega} v' \left(\frac{y^*(\sigma)}{\bar{z}_0} + h^*(\sigma) \right) \right)^- & \text{if} \quad f > f^*(\sigma). \end{cases}$$

(b) For unemployed agents, the subsidy rate is zero: If $\sigma \in Z_0^*$, then $s(\sigma, f) = 0 \; \forall f$.
(c) For all $\sigma \in Z$, the optimal transfer scheme is set as follows:

$$T(\sigma) = y^*(\sigma) - c^*(\sigma) - f^*(\sigma) + s(\sigma, f^*(\sigma)) f^*(\sigma);$$

where $c^*(\cdot)$ and $y^*(\cdot)$ are the consumption and income functions of the second best allocation.

Proof. Consider the following maximisation problem for all $z, \sigma \in Z^2$:

$$\hat{V}(\sigma|z) := \max_f \; y^*(\sigma) - T(\sigma) - (1 - s(\sigma, f)) f - v \left(\frac{y^*(\sigma)}{z} - \frac{\omega - f}{\omega} \right).$$

From the arguments we made in the main text, the piecewise linear function $s(\sigma, f)$ implies that the solution to the above maximization problem is $f^*(\sigma)$ for each $z \notin Z_0$ and for $z = \bar{z}_0$. From the expression of T, when $f = f^*(\sigma)$, we have:

$$y^*(\sigma) - T(\sigma) - (1 - s(\sigma, f^*(\sigma))) f^*(\sigma) = c^*(\sigma).$$

Since declaring σ forces the agent to choose $y^*(\sigma)$, for all such z we have $\hat{V}(\sigma|z) = V^*(\sigma|z)$, the second best value for each declaration σ. That is, each type z, such that either $z \notin Z_0$ or $z = \bar{z}_0$ declaring σ and possibly deviating in f, gets at most $V^*(\sigma|z)$. Since the second best allocation is incentive compatible, we have shown that the proposed scheme is robust to joint deviations in σ and f for all such types. Consider now unemployed agents: $z \in Z_0$ and $z < \bar{z}_0$. Since all unemployed receive the same utility in equilibrium, we have for all these agents $V^*(z|z) = V^*(\bar{z}_0|\bar{z}_0)$. On the other hand, it is immediate to see that for all $z < \bar{z}_0$, $\hat{V}(\sigma|z) \leq \hat{V}(\sigma|\bar{z}_0)$. The fact that agent $\bar{z}_0$ does not want to deviate hence implies that none of these agents want to deviate either. In summary, we have shown that each type $z \in Z$ chooses to tell the truth, produces $y^*(z)$ and spends $f^*(z)$ in formal child care. Since transfers $T(\cdot)$ are adjusted by the child care subsidy to satisfy the government budget constraint, the proof is complete. ∎

A.5 Proof of Proposition 4

Under Assumption 4, there is a $\overline{T} \in \mathbb{R}$ such that the following subsidy rates and transfers implement the constrained optimum.

(a) For employed agents (who earn $y > 0$), we have:

$$s(y,f) = \begin{cases} \left(1 - \frac{1}{\omega}v'\left(\frac{y}{z_N} + 1 - \frac{f(y)}{\omega}\right)\right)^+ & if \quad f \leq f(y)\,; \\ \left(1 - \frac{1}{\omega}v'\left(\frac{y}{\bar{z}_0} + 1 - \frac{f(y)}{\omega}\right)\right)^- & if \quad f > f(y)\,; \end{cases}$$

$$if\ y \in \mathcal{Y}\ then\ T(y) = y - c(y) - f(y) + s(y, f(y))\, f(y)\,;\ otherwise\ T(y) = \overline{T}.$$

(b) For unemployed agents (with $y = 0$), the second best allocation is implemented by having:

$$s(0,f) \equiv 0, \qquad and \qquad T(0) = -c(0) - f(0)\,.$$

Proof. Let $\overline{y}$ be such that $v'\left(\frac{\overline{y}}{z_N}\right) = z_N$ and $\overline{T} := \overline{y} + \max_{(\sigma,z)\in Z^2} V^*(\sigma|z)$. Agent z solves:

$$\max_{y\geq 0, f\leq\omega} \quad y - T(y) - (1 - s(y,f))f^+ - v\left(\frac{y}{z} - \frac{\omega - f}{\omega}\right).$$

Under Assumption 4, for each $y \in \mathbb{R}_+$, there is only one value for $f(y)$ and hence, a well defined subsidy rate schedule $s(y, f)$. Since for each σ there is only one value of $y \in \mathcal{Y}$, to each σ in the direct mechanism there is only one pair of values y and f, those for type $z = \sigma$. Moreover, it is immediate to see that the punishment induced by $\overline{T}$ implies that no agent will ever choose $y \notin \mathcal{Y}$. For $y \in \mathcal{Y}$, however, the agent has weakly less joint deviations available compared to those considered in the implementation of Proposition 3. Moreover, for $y \in \mathcal{Y}$, the welfare and net revenues for the government are as in the direct mechanism. The result hence follows from Proposition 3. ∎

B Selection and Calibration Procedures

B.1 Agent's Private Problem

Before describing the selection and calibration procedures, we first try to understand mothers' private decisions given the existing US tax and benefit system (E). We can rewrite the agent's private problem (9) as:

$$U(z;T,\omega) = \max_{l,h\geq 0} \quad zl - T^E(y,f) - \omega(1-h)^+ - \frac{1}{\theta}\frac{(l+h)^{1+\gamma}}{1+\gamma},$$

where earnings $y = zl$ and formal child care cost $f = \omega(1-h)$. Let $l(z;T,\omega)$ be the labor supply choice for agent z under the transfer program T and child care cost ω. In our framework, the *reservation wage* $R(T,\omega)$ can be identified with the hypothetical type for which $l(z;T,\omega) > 0$ if $z > R(T,\omega)$ and $l(z;T,\omega) = 0$ if $z < R(T,\omega)$. Obviously, we do not observe the distribution of types for wages below $R(T,\omega)$. In order to identify the distribution of z for $z < R(T,\omega)$, in the next section, we assume that (the types of) mother with kids *above* six years of age are distributed as our group of reference. When kids are grown up, mothers plausibly face a lower amount of child care needs. To ease the exposition, suppose that mothers with grown up kids face no child care needs at all. This would correspond to a $\omega' = 0$. It is immediate to see that, since in the US tax scheme $T^E_y(0,0) < 1$ (i.e, income tax is less than 100% at zero income) and when there are no child care needs $f(z;T,0) = 0$ for all z, the reservation wage for these agents is zero, i.e., $R(T,0) = 0$. In general, we will assume that a reduction in child care needs reduces the reservation wage.

B.2 Wage Imputation

Wages for working mothers are computed as yearly gross earnings divided by total hours of work in one year.[26] On the other hand, non working mothers have no earnings. In our model, a mother may not be working either because (i) she has no employment opportunities ($z_1 = 0$) or (ii) her wage is below her reservation wage. As described in the text, we consider the involuntarily unemployed as those with no employment opportunities. We now focus our analysis on the remaining mothers, that is, those who are either working or voluntarily unemployed.

Our log wage function is given by:

$$\ln wage_i = X_i\beta + \epsilon_i,$$

where X_i is a vector of demographic characteristics such as age, education, health, ethnicity and number

[26] We drop mothers with wages above \$40 which consist of 39 observations (approximately 1% of the sample). Those mothers are very sparsely distributed between a wage range of \$40 to \$276.

of children, that are correlated with wages, and ϵ_i is an unobserved component. β is a vector of coefficients that we aim to estimate so that we can impute wages for the unemployed. Note, however, that wages are observed only for the employed. If there is a correlation between wages and the decision to work, the distribution of ϵ_i will be truncated. We therefore would not be able to rely on Ordinary Least Squares (OLS) regressions to estimate β and would need to account for the selection of agents into work.

From the agent's private problem in Section B.1, she works if $z > R(T, \omega)$, where R is the reservation wage. As we just saw, mothers with child care needs have a positive reservation wage. We therefore model the work decision of the agent as:

$$work_i = 1\left[\delta lnwage_i - \gamma K_i - \eta_i > 0\right],$$

where K_i is a dummy variable that captures the child care needs of agents (presence of children aged below 6) and therefore $K_i = 1$ reflects a positive reservation wage ($R > 0$) of mothers. The random variable η_i is an unobserved determinant of work participation that may be present in the real world, and δ and γ are coefficients to be estimated.

Using the equation for log wage, we can rewrite the work decision as:

$$work_i = 1\left[X_i\psi - \gamma K_i - u_i > 0\right],$$

where $u_i = \delta\epsilon_i - \eta_i$. The unobserved terms are assumed to follow a bivariate normal distribution:

$$\begin{bmatrix} \epsilon_i \\ u_i \end{bmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_\varepsilon & \rho \\ \rho & 1 \end{bmatrix}\right),$$

where σ_ε is the variance of ϵ_i and ρ is the correlation between ϵ_i and u_i.

Thus, the conditional mean of log wages is given by:

$$E\left[\ln wage_i | work_i = 1\right] = X_i\beta + \rho\sigma_\epsilon\lambda_i\left(X_i\psi - \gamma K_i\right),$$

where $\lambda_i\left(X_i\psi - \gamma K_i\right) = \frac{\phi(X_i\psi - \gamma K_i)}{\Phi(X_i\psi - \gamma K_i)}$ and ϕ and Φ are the normal pdf and cdf respectively. λ_i is the inverse mills ratio that takes into account the fact that the distribution of ϵ_i is truncated. Note that even in the absence of the unobserved determinant of work, η_i, we would still have a selection issue. This is because in this case, $u_i = \delta\epsilon_i$, so that there would still be a correlation between wages and the work decision.

We use the whole sample of single mothers aged between 18 and 50 and with children under 18 from March 2010 CPS data for our estimation purposes. Table 6 reports summary statistics for this group.

The imputation is done using the Heckman two step estimation procedure (Wooldridge, 2002). First run a probit using work status as the dependent variable and construct the inverse mills ratio. In the second stage, run an OLS regression using log of wages as the dependent variable and controlling for

Table 6: **Summary Statistics for Single Mothers with Children under 18**

Variable	Mean	s.d.	Variable	Mean	s.d.
Age	30.6	9.62	Black	0.21	0.41
High school graduate	0.29	0.49	Proportion working	0.62	0.49
College or university	0.51	0.50	Yearly hours of work (if > 0)	1,559	766
No. of children under 6	0.45	0.69	Wage per hour (if > 0)	13.63	8.31
No. of children under 18	1.67	1.92	Has a child under age 6	0.35	0.48
White	0.70	0.46	No. of observations	7,060	

Source: March 2010 CPS data on single women with at least one child aged below 18. We limit the sample to women who are not involuntarily unemployed but who are either working or voluntarily unemployed (out of the labor force).

demographics X and the inverse mills ratio.

In order to identify our selection correction term, we rely on the non-linearity of the inverse Mills ratio and on the use of an exclusion restriction in the work equation. The exclusion restriction needs to be a variable which may affect mother's work decision but not her wages. We use a dummy variable indicating whether a mother has a child under 6 for this purpose. While the total number of children may be correlated with a women's past work decisions and therefore work experience and wages, once we control for the total number of children, we do not expect the presence of a child under 6 to immediately affect her wages although it may affect her current work decision. This corresponds to our variable K_i which captures child care needs of mothers.

Table 7 reports regression results for our selection and wage equations. As can be seen from our selection equation in column (i), having a child aged below 6 has a negative and statistically significant impact on the work decision of mothers. Moreover, from our wage equation in column (ii), the coefficient of the mills ratio is positive and significant suggesting that individuals who work tend on average to have higher wages.

As discussed in the text, we are interested in the potential wage distribution of voluntarily unemployed mothers who would have been working if they did not have child care needs. We therefore impute their potential log wage as:

$$E\left[\ln wage_i | X_i\psi - \gamma K_i \leq u_i \leq X_i\psi\right] = X_i\beta - \rho\sigma_\epsilon \left[\frac{\phi\left(X_i\psi - \gamma K_i\right) - \phi\left(X_i\psi\right)}{\Phi\left(X_i\psi - \gamma K_i\right) - \Phi\left(X_i\psi\right)}\right].$$

From there, we can infer the potential wage distribution of voluntarily unemployed mothers by finding out the proportion of mothers with a given potential wage.

Table 7: **Selected Coefficients from Work and Wage Regressions**

Dependent variable	(i) work		(ii) lnwage	
	coef	s.e.	coef	s.e.
Age	0.194**	(0.015)	0.148*	(0.021)
High school graduate	1.111**	(0.346)	0.697*	(0.294)
Undergraduate degree	1.731**	(0.352)	1.225**	(0.314)
No. of children under 18	-0.038	(0.056)	-0.022	(0.033)
Fair health	0.788**	(0.118)	0.249†	(0.129)
Good health	1.300**	(0.110)	0.464**	(0.167)
Very good health	1.300**	(0.110)	0.606**	(0.182)
Excellent health	1.490**	(0.111)	0.621**	(0.182)
White	0.251**	(0.065)	0.065	(0.044)
Black	0.173**	(0.075)	0.047	(0.046)
Any child under 6	-0.073*	(0.037)		
Mills			0.529***	(0.177)
No. of observations	7,060			

Standard errors reported in brackets. Controls also include age squared, number of children squared, average unemployment rate in state of residence and state dummies. †significant at 10%, *significant at 5% and **significant at 1%.

B.3 Calibration

US Tax and Benefits System

We base our calibration on the 2010 US tax and benefit system.

Unemployment benefits Unemployment benefits are set at $5,500 such that, given the US tax and benefit system, the proportion of working mothers predicted by our model fit the proportion of working mothers (56%) in our CPS sample. Since families with two children receive on average $412 TANF benefits per month (US Department of Health and Human Services, 2011), we interpret unemployment benefits as the sum of yearly TANF benefits of $4,944 and of additional benefits of $556 which may constitute of unemployment insurance benefits or food stamps that an unemployed individual may be eligible for. We do not explicitly set unemployment insurance benefits as young mothers may not be eligible for them if they have no previous work experience.

Federal taxes Taxable income is based on earnings minus standard deductions of $5,700 for a single childless person and of $8,400 for the head of household. In addition, each taxpayer and dependent get personal exemptions of $3,650. The tax rates vary according to income levels as follows (Taxes About):

Tax rate	Taxable income
10%	Less than $8,375
15%	$8,375 - $34,000
25%	$34,000 - $82,400
28%	$82,400 - $171,850
33%	$171,850 - $373,650
35%	$373,650 and above

Social Security taxes The Social Security base wage was $106,800 in 2010 and the employee rate 7.65% (Payroll Experts).

Earned Income Tax Credit The EITC is a refundable tax credit payable to working families. Earned income must be below $40,363 for a single parent with two children aged below 18 and the maximum credit was $5,036 in 2010 (Tax Policy Center, 2010a). The phase-in rate was 40% and the phase-out rate 21.06% while the phase-out income range starts at $16,420. In the phase-in income range, EITC benefits are computed as 40% of earned income up to the maximum credit of $5,036. In the phase-out income range, EITC benefits are the difference between the maximum credit and 21.06% of income earned above $16,420. For a single childless individual, EITC benefis are phased-in at a rate of 7.65% up to a maximum of $457. Benefits are subsequently phased-out at a rate of 7.65% until earnings of $13,460 beyond which EITC benefits are zero.

Dependent Care Tax Credit The dependent care tax credit is a non-refundable tax credit as described in Section 2. It covers 35% of cost of formal child care up to a cap of $6k for two children families earning less than $15,000. The tax credit rate declines by 1% for each $2,000 additional income until it reaches a constant rate of 20% for families with annual gross income above $43,000.

Child Care and Development Fund We set the CCDF rate to 90% which is the recommended subsidy rate under Federal guidelines. We take into account the fact that only a certain proportion of eligible households received the CCDF subsidy: 39% of potentially eligible children living in households below the poverty threshold, 24% of potentially eligible living in households with income between 101 to 150% of the poverty threshold, and 5% of potentially eligible children living in household with income above 150% of the poverty threshold but below the CCDF eligibility threshold of 85% of state median income (US Department of Health and Human Services, 2009). We therefore compute the average CCDF subsidy rate as 35.1%, 21.6%, and 4.5% for households with income below, between 101% and 150%, and above 150% of the poverty threshold, respectively. The poverty threshold for a single parent with two children was $17,568 and US median earnings was $32,349 in 2010.

Calibration of θ

The calibration of θ is done as follows: define a grid over $\theta \in [0.5, 2.5]$ with equally spaced intervals of 0.1. For each θ and z, we find the optimal labor supply predicted by our model $l(\theta, z)$ given the actual US tax and benefit system. Given the selection corrected empirical distribution of wages $\pi(z)$, we then compute the average labor supply predicted by our model for each θ, that is, we compute $\bar{l}(\theta) = \sum_z \pi(z)\, l(\theta, z)$. We find θ by minimizing the square of the distance between average labor supply predicted in our model and average labor supply in the data $\bar{l}_{data} = \sum_z \pi(z)\, l_{data}(z)$:

$$\hat{\theta} = argmin\left[\bar{l}(\theta) - \bar{l}_{data}\right]^2$$

After obtaining $\hat{\theta}$, we define a finer grid over θ (within a smaller interval that is inclusive of $\hat{\theta}$) with equally spaced intervals of 0.01 and repeat the procedure in order to get a more precise estimate of θ.

Calibration of M

Net transfers that the US government already allocates to mothers are computed as follows. Given our calibrated values of θ, we simulate the chosen allocations $(c(z), y(z), f(z))$ for each type z, given the actual tax and benefit system. Based on the computed earnings $y(z)$, we then compute the Federal and SS taxes, EITC, DCTC and CCDF benefits as described above. Unemployed mothers receive unemployment benefits inclusive of TANF. Given the different net transfers received by each z, we take the average:

$$M = -\sum_z \pi(z)\, T^E(y(z), f(z))$$

where $T^E(y(z), f(z))$ are the net taxes computed based on the actual US tax and benefit system.

B.4 Numerical Algorithm

We numerically solve for the constrained optimal allocations using Matlab. First, we impose non-negativity constraints on h, y. We then solve for the government problem using the following steps:

1. *Relaxed problem*. (a) Make an initial guess of values for the optimal allocations (c, y, h) and maximize welfare by imposing the government budget constraint with equality and the LDIC with inequality. (b) Use the solution in (a) as the new initial guess of values for (c, y, h) and maximize welfare by imposing the government budget constraint with equality and all the DIC with inequality. We have 1,275 DIC in total.

2. *Ex-post verification*. After having obtained the solution for the relaxed problem in point 1, we check whether all the UIC are satisfied. If all LDIC are binding we do not need to check the UIC as they

are automatically satisfied from the proof of Lemma 1. Similarly, in the Ralwsian case, Lemma 1 guarantees this ex-post check step is not needed.

3. *Full problem*. If the conditions in Step 2 are not satisfied, then use the solution in Step 1(b) as the new initial guess of values for (c, y, h) and maximize welfare by imposing the government budget constraint with equality and all the DIC and UIC with inequality. That is, solve the full-blown problem. Excluding the 50 UIC for z_1, we have $2,500$ incentive constraints in total.

B.5 Further Sensitivity Analysis

We conduct further sensitivity analysis with respect to the efficient child allowances and child care subsidy rates. The implementation is as in Section 6.2, where we keep US net taxes (federal and SS taxes net of EITC if employed and TANF if unemployed) fixed.

Figure 8 illustrates the baseline and US optimal child allowances and child care subsidy rates as in Figure 6. Recall that the number of wage bins as per our calibration in Section 6 is 50, which implies a maximum wage of \$32.21. As the optimal child care subsidy rate relies on z_N, we vary the number of wage bins to see how sensitive the rate is to the maximum wage. Wage bin grids of 25, 100, and 500 points yield z_N equal to \$28.80, \$35.60, and \$38.80, respectively. As can be seen from Figure 8, child allowances and subsidy rates for the different wage bins follow closely the baseline optimal ones.

Sensitivity Analysis on Child Care Subsidy Rates and Child Allowances

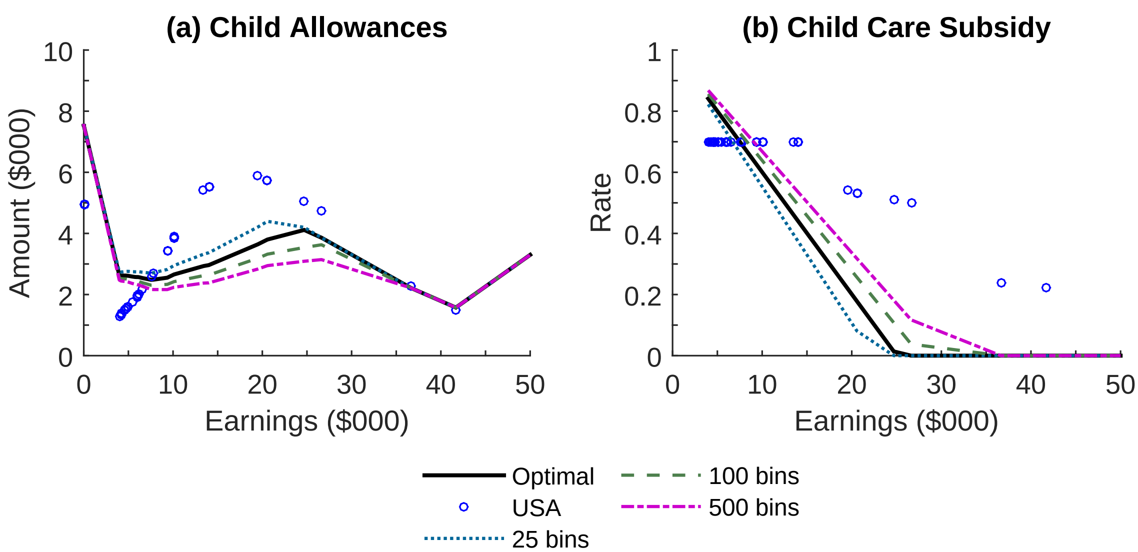

Figure 8: Panel (a) illustrates child allowances and Panel (b) illustrates child care subsidy rates. The US net income taxes are kept fixed (federal and SS taxes net of EITC if employed and TANF if unemployed) as in Figure 6. The rates and allowances are illustrated for the baseline optimal scheme, for the US, for the baseline optimal scheme when using 25, 100 and 500 wage bins.